```
                      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4195> which encodes the amino acid sequence <SEQ ID 4196>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1985(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 28 SFGTIRNSTALKQLTLDSLNLLSFGTIRNSTALK 61
          SFGTI+NS ALKQ    + +N  SFGTI+NS ALK
Sbjct: 7  SFGTIQNSIALKQKAQEEINQRSFGTIQNSIALK 40
 Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 6  SFGTIRNSTALKLYAKQSPAFRSFGTIRNSTALK 39
          SFGTI+NS ALK  A++    RSFGTI+NS ALK
Sbjct: 7  SFGTIQNSIALKQKAQEEINQRSFGTIQNSIALK 40
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1372

A DNA sequence (GBSx1457) was identified in *S.agalactiae* <SEQ ID 4197> which encodes the amino acid sequence <SEQ ID 4198>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1407(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4199> which encodes the amino acid sequence <SEQ ID 4200>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 154/221 (69%), Positives = 187/221 (83%)

Query: 1    MIKINFPILDEPLVLSNATILTIEDVSVYSSLVKHFYQYDVDEHLKLFDDKQKSLKATEL 60
            ++ +NF +LDEP+ L    TIL +EDV V+S +V++ YQY+ D  LK FD K K++K +E+
```

```
Sbjct: 8   LMNLNFSLLDEPIPLRGGTILVLEDVCVFSKIVQYCYQYEEDSELKFFDHKMKTIKESEI 67

Query: 61  MLVTDILGYDVNSAPILKLIHGDLENQFNEKPEVKSMVEKLAATITELIAFECLENELDL 120
           MLVTDILG+DVNS+ ILKLIH DLE+QFNEKPEVKSM++KL ATITELI FECLENELDL
Sbjct: 68  MLVTDILGFDVNSSTILKLIHADLESQFNEKPEVKSMIDKLVATITELIVFECLENELDL 127

Query: 121 EYDEIKILELIKALGVKIETQSDTIFEKCFEIIQVYHYLTKKNLLVFVNSGAYLTKDEVI 180
           EYDEI ILELIK+LGVK+ETQSDTIFEKC EI+Q++ YLTKK LL+FVNSGA+LTKDEV
Sbjct: 128 EYDEITILELIKSLGVKVETQSDTIFEKCLEILQIFKYLTKKKLLIFVNSGAFLTKDEVA 187

Query: 181 KLCEYINLMQKSVLFLEPRRLYDLPQYVIDKDYFLIGENMV 221
            L EYI+L   +VLFLEPR LYD PQY++D+DYFLI +NMV
Sbjct: 188 SLQEYISLTNLTVLFLEPRELYDFPQYILDEDYFLITKNMV 228
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1373

A DNA sequence (GBSx1458) was identified in *S.agalactiae* <SEQ ID 4201> which encodes the amino acid sequence <SEQ ID 4202>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0842(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9783> which encodes amino acid sequence <SEQ ID 9784> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB83918 GB:AL162753 hypothetical protein NMA0629 [Neisseria
          meningitidis Z2491]
 Identities = 45/104 (43%), Positives = 65/104 (62%), Gaps = 2/104 (1%)

Query: 4   RYMRMILMFDMPTETAEERKAYRKFRKFLLSEGFIMHQFSVYSKLLLNNTANNAMIGRLK 63
           ++MR+I+ FD+P  TA +RKA  +FR+FLL +G+ M Q SVYS+++    +      RL
Sbjct: 5   KFMRIIVFFDLPVITAAKRKAANQFRQFLLKDGYQMLQLSVYSRIVKGRDSLQKHHNRLC 64

Query: 64  VNNPKKGNITLLTVTEKQFARMVYLHGERNT--SVANSDSRLVF 105
            N P++G+I  L +TEKQ+A M  L GE  T     NSD  L+F
Sbjct: 65  ANLPQEGSIRCLEITEKQYAAMKLLLGELKTQEKKVNSDQLLLF 108
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4203> which encodes the amino acid sequence <SEQ ID 4204>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0822(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 97/112 (86%), Positives = 107/112 (94%)

Query: 1   MSYRYMRMILMFDMPTETAEERKAYRKFRKFLLSEGFIMHQFSVYSKLLLNNTANNAMIG 60
           MSYRYMRMILMFDMPT+TAEERKAYRKFRKFLLSEGFIMHQFS+YSKLLLNNTANNAMIG
```

```
Sbjct: 1   MSYRYMRMILMFDMPTDTAEERKAYRKFRKFLLSEGFIMHQFSIYSKLLLNNTANNAMIG 60

Query: 61  RLKVNNPKKGNITLLTVTEKQFARMVYLHGERNTSVANSDSRLVFLGDSYDQ 112
           RL+ +NP KGNITLLTVTEKQFARM+YLHGERN  +ANSD RLVFLG+++D+
Sbjct: 61  RLREHNPNKGNITLLTVTEKQFARMIYLHGERNNCIANSDERLVFLGEAFDE 112
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1374

A DNA sequence (GBSx1459) was identified in *S.agalactiae* <SEQ ID 4205> which encodes the amino acid sequence <SEQ ID 4206>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3185(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB83919 GB:AL162753 hypothetical protein NMA0630 [Neisseria
           meningitidis Z2491]
 Identities = 71/224 (31%), Positives = 122/224 (53%)

Query: 4   WRTVVVNTHSKLSYKNNHLIFKDSYQTEMIHLSEIDILIMETTDIVLSTMLIKRLVDENI 63
           WR++++    KLS +   L+ + + ++  + L +I ++I+E  + +++  L+  L +
Sbjct: 3   WRSLLIQNGGKLSLQRRQLLIQQNGESHTVPLEDIAVIIIENRETLITAPLLSALAEHGA 62

Query: 64  LVIFCDDKRLPTAMLMPYYARHDSSLQLSRQMSWIEDVKADVWTSIIAQKILNQSFYLGE 123
            ++ CD++ LP    +PY   H     L  Q++  E +K  +W  I+ QKILNQ+F   E
Sbjct: 63  TLLTCDEQFLPCGQWLPYAQYHRQLKILKLQLNISEPLKKQLWQHIVRQKILNQAFVADE 122

Query: 124 CSFFEKSQSIMNLYHDLEPFDPSNREGHAARIYFNTLFGNDFSREQDNPINAGLDYGYSL 183
                 ++ +  L  ++   D  NRE  AA +YF  LFG  F+R  +N +NA L+Y Y++
Sbjct: 123 TGNDLAAKRLRTLASEVRSGDTGNREAQAAALYFQALFGEKFTRNDNNAVNAALNYTYAV 182

Query: 184 LLSMFAREVVKCGCMTQFGLKHANQFNQFNLASDIMEPFRPIVD 227
           L +  AR +   G +   GL H ++ N FNLA D +EP RP+ D
Sbjct: 183 LRAAVARALTLYGWLPALGLFHRSELNPFNLADDFIEPLRPLAD 226
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4207> which encodes the amino acid sequence <SEQ ID 4208>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3185(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 239/289 (82%), Positives = 271/289 (93%)

Query: 1   MAGWRTVVVNTHSKLSYKNNHLIFKDSYQTEMIHLSEIDILIMETTDIVLSTMLIKRLVD 60
           MAGWRTVVVNTHSKLSYKNNHLIFKD+Y+TE+IHLSEIDIL++ETTDIVLSTML+KRLVD
Sbjct: 1   MAGWRTVVVNTHSKLSYKNNHLIFKDAYKTELIHLSEIDILLLETTDIVLSTMLVKRLVD 60

Query: 61  ENILVIFCDDKRLPTAMLMPYYARHDSSLQLSRQMSWIEDVKADVWTSIIAQKILNQSFY 120
           EN+LVIFCDDKRLPTAMLMP+Y RHDSSLQL +QMSW E VK+ VWT+IIAQKILNQS Y
```

```
Sbjct: 61  ENVLVIFCDDKRLPTAMLMPFYGRHDSSLQLGKQMSWSETVKSQVWTTIIAQKILNQSCY 120

Query: 121 LGECSFFEKSQSIMNLYHDLEPFDPSNREGHAARIYFNTLFGNDFSREQDNPINAGLDYG 180
           LG CS+FEKSQSIM+LYH LE FDPSNREGHAARIYFNTLFGNDFSR+ ++PINAGLDYG
Sbjct: 121 LGACSYFEKSQSIMDLYHGLENFDPSNREGHAARIYFNTLFGNDFSRDLEHPINAGLDYG 180

Query: 181 YSLLLSMFAREVVKCGCMTQFGLKHANQFNQFNLASDIMEPFRPIVDRIIYENRQSDFVK 240
           Y+LLLSMFAREVV  GCMTQFGLKHANQFNQFN ASDIMEPFRP+VD+I+YENR   F K
Sbjct: 181 YTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNFASDIMEPFRPLVDKIVYENRNQPFPK 240

Query: 241 MKRELFSMFSETYSYNGKEMYLSNIVSDYTKKVIKSLNSDGNGIPEFRI 289
           +KRELF++FS+T+SYNGKEMYL+NI+SDYTKKV+K+LN++G G+PEFRI
Sbjct: 241 IKRELFTLFSDTFSYNGKEMYLTNIISDYTKKVVKALNNEGKGVPEFRI 289
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1375

A DNA sequence (GBSx1460) was identified in *S.agalactiae* <SEQ ID 4209> which encodes the amino acid sequence <SEQ ID 4210>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1109(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB73943 GB:AL139078 hyopthetical protein Cj1523c [Campylobacter
            jejuni]
 Identities = 165/746 (22%), Positives = 291/746 (38%), Gaps = 115/746 (15%)

Query: 318  LSASMIQRYDEHREDLKQLKQFVKASLPEKYQEI--FADSSKDGYAGYIEGKTNQEAFYK 375
            L+ S  +R    +  L  LK  +       Y++   F +S    Y G +      E  ++
Sbjct: 50   LARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISP--YELRFR 107

Query: 376  YLSKLLTKQEDSENFLE--KIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQS----- 428
             L++LL+KQ+ +   L   K +  D ++     + G+I   +   E K +   QS
Sbjct: 108  ALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEK-LANYQSVGEYL 166

Query: 429  --EYYPFLKENQDRIEKILTFRIPYY-----------IGPLAREKSDFAW-MTRKTDDSI 474
              EY+   KEN      +   +  Y            +  + +++ +F +  ++K ++ +
Sbjct: 167  YKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEV 226

Query: 475  RPWNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKV--RYKN 532
                F      +++ + F H + N  F+  +EK  PK+S +   F     +  +    KN
Sbjct: 227  LSVAFY-----KRALKDFSHLVGNCSFFT-DEKRAPKNSPLAFMFVALTRIINLLNNLKN 280

Query: 533  EQGETYFFDSNIKQEIFDGVFKEHRKVSK--KKLLDFLAKEYEEFRIVDVIGLDKENKAF 590
             +G  Y  D      + + V K      K  KKLL  L+ +YE            E   +
Sbjct: 281  TEGILYTKDD--LNALLNEVLKNGTLTYKQTKKLLG-LSDDYE---------FKGEKGTY 328

Query: 591  NASLGTYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQL 650
                  Y +  K L +  L   D    L +I + +TL +D   +KK L  Y     ++Q+
Sbjct: 329  FIEFKKYKEFIKALGEHNLSQDD----LNEIAKDITLIKDEIKLKKALAKYD--LNQNQI 382

Query: 651  KKLYRRHYTGWGRLSAKLINGIRDK--ESQKTILDYLIDDGRSNRNFMQLINDDGLSFKS 708
              L +  +     +S K +  +     E +K       D+  +  N    IN+D   F
Sbjct: 383  DSLSKLEFKDHLNISFKALKLVTPLMLEGKK------YDEACNELNLKVAINEDKKDFLP 436

Query: 709  IISKAQAGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMAREN 768
              ++        N           P + + I +  K+++ L+K  G +  +I +E+ARE
```

```
Sbjct: 437  AFNETYYKDEVTN-----------PVVLRAIKEYRKVLNALLKKYG-KVHKINIELAREV 484

Query: 769  QTTNQGR----RNSRQRYKLLDDG---VKNLASDLNG-NILKEYPTDNQALQNERLFLYY 820
                + R   +  + YK   D     + L   +N  NILK             L L+
Sbjct: 485  GKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILK-------------LRLFK 531

Query: 821  LQNGRDMYTGEALDIDNLSQ---YDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPS 877
             Q     Y+GE + I +L      +IDHI P +   DDS  N+VLV + +N+ K +  P
Sbjct: 532  EQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTP- 590

Query: 878  LEIVKDCKVFWKKL--LDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKH 935
             E   +    W+K+  L   L ++++   L K         ++  F  R L +TR I +
Sbjct: 591  FEAFGNDSAKWQKIEVLAKNLPTKKQKRILDK----NYKDKEQKNFKDRNLNDTRYIARL 646

Query: 936  VARI---------LDERFNNELDSKGRRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNY 986
            V            L+   N +L+   ++  KV +      L S  R  +GF      N+
Sbjct: 647  VLNYTKDYLDFLPLSDDENTKLNDT-QKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHL 705

Query: 987  HHAHDAYLNAVVAKAILTKYPQLEPE 1012
            HHA DA + A    +I+  +   + E
Sbjct: 706  HHAIDAVIIAYANNSIVKAFSDFKKE 731
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4211> which encodes the amino acid sequence <SEQ ID 4212>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0973(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 881/1380 (63%), Positives = 1088/1380 (78%), Gaps = 22/1380 (1%)

Query: 1    MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAA 60
            M+K YSIGLDIGTNSVGW++ITD+YKVP+KK +VLGNTD+  IKKNLIGALLFD G TA
Sbjct: 1    MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE 60

Query: 61   DRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIFA 120
              RLKRTARRRYTRR+NRI YLQEIF+ EM+KVDDSFFHRLE+SFLVEEDK+  ++PIF
Sbjct: 61   ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG 120

Query: 121  TLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNT 180
             + +E  YHEK+ TIYHLRK+L D  +KADLRLIY+ALAH+IKFRGHFLIE D  +  N+
Sbjct: 121  NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD-LNPDNS 179

Query: 181  DISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGIFA 240
            D+ K +   ++ +N  FE N + +  VD +AIL+ ++SKS + + ++AQ P +K  G+F
Sbjct: 180  DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG 239

Query: 241  EFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDS 300
              + L +G   +FK  F+L +   LQ +KD+YD+DL+NLL QIGD++ADLF AAK L D+
Sbjct: 240  NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA 299

Query: 301  VLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGY 360
            +LLS IL V    TKAPLSASMI+RYDEH +DL  LK  V+  LPEKY+EIF D SK+GY
Sbjct: 300  ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY 359

Query: 361  AGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTEL 420
            AGYI+G  +QE FYK++  +L K + +E  L K+  ED LRKQRTFDNGSIPHQ+HL EL
Sbjct: 360  AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL 419

Query: 421  KAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDDSIRPWNFE 480
             AI+RRQ ++YPFLK+N+++IEKILTFRIPYY+GPLAR  S FAWMTRK++++I PWNFE
```

```
Sbjct: 420  HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE 479

Query: 481  DLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNE-QGETYF 539
            ++VDK  SA++FI RMTN D  LP EKVLPKHSL+YE FTVYNELTKV+Y  E   +  F
Sbjct: 480  EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF 539

Query: 540  FDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLGTYHD 599
                 K+ I D +FK +RKV+ K+L +   K+ E F  V++ G++     FNASLGTYHD
Sbjct: 540  LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLGTYHD 596

Query: 600  LEKIL-DKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHY 658
            L KI+ DKDFLDN +NE ILEDIV TLTLFEDREMI++RL+ Y  LF +  +K+L RR Y
Sbjct: 597  LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY 656

Query: 659  TGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQAGSH 718
            TGWGRLS KLINGIRDK+S KTILD+L  DG +NRNFMQLI+DD L+FK  I KAQ
Sbjct: 657  TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ 716

Query: 719  SDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMG-YEPEQIVVEMARENQTTNQGRRN 777
             D+L E +  LAGSPAIKKGILQ++K+VDELVKVMG ++PE IV+EMARENQTT +G++N
Sbjct: 717  GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN 776

Query: 778  SRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEALDIDN 837
            SR+R K +++G+K L S    ILKE+P +N  LQNE+L+LYYLQNGRDMY  + LDI+
Sbjct: 777  SRERMKRIEEGIKELGS----QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR 832

Query: 838  LSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLM 897
            LS YD+DHI+PQ+F+KDDSIDN+VL  S KNRGKSD+VPS E+VK  K +W++LL+AKL+
Sbjct: 833  LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI 892

Query: 898  SQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKGRRIRK 957
            +QRK+DNLTKAERGGL+  DKA FI+RQLVETRQITKHVA+ILD R N + D   + IR+
Sbjct: 893  TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE 952

Query: 958  VKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGD 1017
            VK++TLKS LVS+FRK+F FYK+RE+NNYHHAHDAYLNAVV  A++ KYP+LE EFVYGD
Sbjct: 953  VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD 1012

Query: 1018 YPKYN-------SYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEI 1070
            Y  Y+       S +    AT K FFYSNIMNFFKT++TLA+G +   +  IE N +TGEI
Sbjct: 1013 YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI 1072

Query: 1071 VWDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPK 1130
            VWDK + FATVRKVLS PQ NIVKKTE+QTGGFSKESIL   NSDKLI RK KD   DPK
Sbjct: 1073 VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK-KD--WDPK 1129

Query: 1131 KYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN 1190
            KYGGFDSP VAYSVLVVA ++KGK++KLK+V ELLGITIMERS FEKNP  FLE+KGY
Sbjct: 1130 KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE 1189

Query: 1191 IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKP 1250
            ++ D +I LPKYSLFELENGR+R+LASAGELQKGNELALP++++ FLYLAS Y + KG P
Sbjct: 1190 VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP 1249

Query: 1251 EEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELANN 1310
            E+ E+KQ FV QH  Y D+I++ I++FSKRVILADANL+K+    Y  +++   + E A N
Sbjct: 1250 EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK-PIREQAEN 1308

Query: 1311 IINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED 1370
            II+LFT T+LGAPAAFK+FD  +DRKRYTSTKEVL++TLIHQSITGLYETRIDL +LG D
Sbjct: 1309 IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD 1368
```

SEQ ID 4210 (GBS317) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 2; MW 179.3kDa) and in Figure 159 (lane 5 & 6; MW 180kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 3; MW 154.3kDa) and in Figure 159 (lane 9 & 10; MW 154kDa).

GBS317-GST was purified as shown in Figure 224, lane 9-10. GBS317-His was purified as shown in Figure 222, lane 9.

GBS317N was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 2-4; MW 116kDa).

GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 6-8; MW 92kDa).

GBS317dN was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 7; MW 116kDa). Purified GBS317dN-GST is shown in Figure 245, lane 8.

GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 188 (lane 13; MW 92kDa). Purified GBS317dC-GST is shown in Figure 245, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1376

A DNA sequence (GBSx1461) was identified in *S.agalactiae* <SEQ ID 4213> which encodes the amino acid sequence <SEQ ID 4214>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.94   Transmembrane  132 - 148 ( 123 - 156)
   INTEGRAL    Likelihood =-11.09   Transmembrane  190 - 206 ( 183 - 209)
   INTEGRAL    Likelihood = -4.94   Transmembrane   95 - 111 (  94 - 115)

----- Final Results -----
              bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related sequence was also identified in GAS <SEQ ID 9133> which encodes the amino acid sequence <SEQ ID 9134>. Analysis of this protein sequence reveals the following:

```
     Possible site: 22

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.32   Transmembrane  126 - 142
   INTEGRAL    Likelihood = -6.90   Transmembrane  178 - 194

----- Final Results -----
              bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 94/204 (46%), Positives = 139/204 (68%)

Query: 5    LMKDKLLVVLTWIWIISLATLATIYIAWLIYPIEIQFLKLEKVVYLKAETIYYNFNKLMI 64
            +M +   ++ +W+W+++LA L TIY  WL YP+E+  LKLE+VV++   + I +N+N L+
Sbjct: 4    VMVENTKLLCSWVWLLALAILITIYSTWLWYPLEVDHLKLEQVVFMSKDAILHNYNGLLN 63

Query: 65   YLTHPFISDLNMPSFPSSEDGLKHFADVKYLFTLAHGLFVILTFPVIYFLRRGWKQKSIF 124
            YLT+PF++ L   +F SS DGLKHFADVK+LF L   +F+ L +P +     +  K K  +
Sbjct: 64   YLTNPFVTRLEFANFHSSADGLKHFADVKWLFHLTQVVFLGLLYPTLKTFTQRLKTKRFW 123
```

```
Query: 125 LYEGFFKIAIMLPIFIVVCAFLLGFDQFFTLFHEVLFPGDSTWQFNPLTDPVIWILPETF 184
           L +     +A + P+ I + A  +GF+ FFTLFH+VLF GDS+W F+PL D VIWILPE F
Sbjct: 124 LLQKPLILAALFPLMIGLMASFIGFEHFFTLFHQVLFVGDSSWLFDPLKDSVIWILPEVF 183

Query: 185 FLHCFIIFLLIYETITIILLIIGR 208
           FLHCF+ F+++YE I   L+ + R
Sbjct: 184 FLHCFLFFMIVYEIILWSLVGLAR 207
```

SEQ ID 4214 (GBS167) was expressed in and purified from *E.coli.* The purified protein is shown in lanes 5 & 6 of Figure 223.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1377

A DNA sequence (GBSx1462) was identified in *S.agalactiae* <SEQ ID 4217> which encodes the amino acid sequence <SEQ ID 4218>. This protein is predicted to be p-nitrophenyl phosphatase (pho2). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3925(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15219 GB:Z99120 similar to N-acetyl-glucosamine catabolism
          [Bacillus subtilis]
 Identities = 121/249 (48%), Positives = 172/249 (68%)

Query: 3   YKGYLIDLDGTIYKGKSRIPAGERFIERLQEKGIPYMLVTNNTTRTPESVQEMLRGFNVE 62
           YKGYLIDLDGT+Y G  +I     F+  L+++G+PY+ VTNN++RTP+ V + L  F++
Sbjct: 4   YKGYLIDLDGTMYNGTEKIEEACEFVRTLKDRGVPYLFVTNNSSRTPKQVADKLVSFDIP 63

Query: 63  TPLETIYTATMATVDYMNDMNRGKTAYVIGEEGLKKAIADAGYVEDTKNPAYVVVGLDWN 122
              E ++T +MAT  ++    +  + YVIGEEG+++AI + G     +N  +VVVG+D +
Sbjct: 64  ATEEQVFTTSMATAQHIAQQKKDASVYVIGEEGIRQAIEENGLTFGGENADFVVVGIDRS 123

Query: 123 VTYDKLATATLAIQNGALFIGTNPDLNIPTERGLLPGAGSLNALLEAATRIKPVFIGKPN 182
           +TY+K A   LAI+NGA FI TN D+ IPTERGLLPG GSL ++L  +T ++PVFIGKP
Sbjct: 124 ITYEKFAVGCLAIRNGARFISTNGDIAIPTERGLLPGNGSLTSVLTVSTGVQPVFIGKPE 183

Query: 183 AIIMNKALEILNIPRNQAVMVGDNYLTDIMAGINNDIDTLLVTTGFTTVEEVPDLPIQPS 242
           +IIM +A+ +L    ++ +MVGDNY TDIMAGIN  +DTLLV TG T  E + D   +P+
Sbjct: 184 SIIMEQAMRVLGTDVSETLMVGDNYATDIMAGINAGMDTLLVHTGVTKREHMTDDMEKPT 243

Query: 243 YVLASLDEW 251
           + + SL EW
Sbjct: 244 HAIDSLTEW 252
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4219> which encodes the amino acid sequence <SEQ ID 4220>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53    Transmembrane  128 - 144 ( 128 - 144)

----- Final Results -----
               bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
```

```
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15219 GB:Z99120 similar to N-acetyl-glucosamine catabolism
           [Bacillus subtilis]
 Identities = 121/250 (48%), Positives = 166/250 (66%), Gaps = 1/250 (0%)

Query: 3   YKGYLIDLDGTIYQGKNRIPAGERFIKRLQERGIPYLLVTNNTTRTPEMVQSMLANQFHV 62
           YKGYLIDLDGT+Y G  +I     F++ L++RG+PYL VTNN++RTP+ V   L + F +
Sbjct: 4   YKGYLIDLDGTMYNGTEKIEEACEFVRTLKDRGVPYLFVTNNSSRTPKQVADKLVS-FDI 62

Query: 63  ETSIETIYTATMATVDYMNDMNRGKTAYVIGETGLKSAIAAAGYVEELENPAYVVVGLDS 122
             + E ++T +MAT  ++    +  + YVIGE G++ AI   G     EN  +VVVG+D
Sbjct: 63  PATEEQVFTTSMATAQHIAQQKKDASVYVIGEEGIRQAIEENGLTFGGENADFVVVGIDR 122

Query: 123 QVTYEMLAIATLAIQKGALFIGTNPDLNIPTERGLMPGAGALNALLEAATRVKPVFIGKP 182
            +TYE  A+  LAI+ GA FI TN D+ IPTERGL+PG G+L ++L  +T V+PVFIGKP
Sbjct: 123 SITYEKFAVGCLAIRNGARFISTNGDIAIPTERGLLPGNGSLTSVLTVSTGVQPVFIGKP 182

Query: 183 NAIIMNKSLEVLGIQRSEAVMVGDNYLTDIMAGIQNDIATILVTTGFTRPEEVPTLPIQP 242
            +IIM +++ VLG   SE +MVGDNY TDIMAGI   + T+LV TG T+ E +     +P
Sbjct: 183 ESIIMEQAMRVLGTDVSETLMVGDNYATDIMAGINAGMDTLLVHTGVTKREHMTDDMEKP 242

Query: 243 DHVLSSLDEW 252
            H + SL EW
Sbjct: 243 THAIDSLTEW 252
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 207/250 (82%), Positives = 227/250 (90%), Gaps = 1/250 (0%)

Query: 3   YKGYLIDLDGTIYKGKSRIPAGERFIERLQEKGIPYMLVTNNTTRTPESVQEMLRG-FNV 61
           YKGYLIDLDGTIY+GK+RIPAGERFI+RLQE+GIPY+LVTNNTTRTPE VQ ML   F+V
Sbjct: 3   YKGYLIDLDGTIYQGKNRIPAGERFIKRLQERGIPYLLVTNNTTRTPEMVQSMLANQFHV 62

Query: 62  ETPLETIYTATMATVDYMNDMNRGKTAYVIGEEGLKKAIADAGYVEDTKNPAYVVVGLDW 121
           ET +ETIYTATMATVDYMNDMNRGKTAYVIGE GLK AIA AGYVE+ +NPAYVVVGLD
Sbjct: 63  ETSIETIYTATMATVDYMNDMNRGKTAYVIGETGLKSAIAAAGYVEELENPAYVVVGLDS 122

Query: 122 NVTYDKLATATLAIQNGALFIGTNPDLNIPTERGLLPGAGSLNALLEAATRIKPVFIGKP 181
            VTY+ LA ATLAIQ GALFIGTNPDLNIPTERGL+PGAG+LNALLEAATR+KPVFIGKP
Sbjct: 123 QVTYEMLAIATLAIQKGALFIGTNPDLNIPTERGLMPGAGALNALLEAATRVKPVFIGKP 182

Query: 182 NAIIMNKALEILNIPRNQAVMVGDNYLTDIMAGINNDIDTLLVTTGFTTVEEVPDLPIQP 241
           NAIIMNK+LE+L I R++AVMVGDNYLTDIMAGI NDI T+LVTTGFT  EEVP LPIQP
Sbjct: 183 NAIIMNKSLEVLGIQRSEAVMVGDNYLTDIMAGIQNDIATILVTTGFTRPEEVPTLPIQP 242

Query: 242 SYVLASLDEW 251
            +VL+SLDEW
Sbjct: 243 DHVLSSLDEW 252
```

A similar DNA sequence was identified in *S.pyogenes* <SEQ ID 4215> which encodes amino acid sequence <SEQ ID 4216>. An alignment of the GAS and GBS sequences follows:

```
Identities = 94/204 (46%), Positives = 139/204 (68%)

Query: 4   VMVENTKLLCSWVWLLALAILITIYSTWLWYPLEVDHLKLEQVVFMSKDAILHNYNGLLN 63
           +M +   ++ +W+W+++LA L TIY  WL YP+E+  LKLE+VV++  + I +N+N L+
Sbjct: 5   LMKDKLLVVLTWIWIISLATLATIYIAWLIYPIEIQFLKLEKVVYLKAETIYYNFNKLMI 64

Query: 64  YLTNPFVTRLEFANFHSSADGLKHFADVKWLFHLTQVVFLGLLYPTLKTFTQRLKTKRFW 123
           YLT+PF++ L   +F SS DGLKHFADVK+LF L   +F+ L +P +    +  K K  +
Sbjct: 65  YLTHPFISDLNMPSFPSSEDGLKHFADVKYLFTLAHGLFVILTFPVIYFLRRGWKQKSIF 124

Query: 124 LLQKPLILAALFPLMIGLMASFIGFEHFFTLFHQVLFVGDSSWLFDPLKDSVIWILPEVF 183
```

```
           L +    +A + P+ I + A  +GF+ FFTLFH+VLF GDS+W F+PL D VIWILPE F
Sbjct: 125 LYEGFFKIAIMLPIFIVVCAFLLGFDQFFTLFHEVLFPGDSTWQFNPLTDPVIWILPETF 184

Query: 184 FLHCFLFFMIVYEIILWSLVGLAR 207
           FLHCF+ F+++YE I   L+ + R
Sbjct: 185 FLHCFIIFLLIYETITIILLIIGR 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1378

A DNA sequence (GBSx1463) was identified in *S.agalactiae* <SEQ ID 4221> which encodes the amino acid sequence <SEQ ID 4222>. This protein is predicted to be oleoyl-acyl carrier protein thioesterase. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3332(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB02069 GB:AB026647 acyl carrier protein thioesterase
           [Arabidopsis thaliana]
 Identities = 59/248 (23%), Positives = 104/248 (41%), Gaps = 30/248 (12%)

Query: 2   GLLYRETYEVPFYESDTNHYMKLPQLLALALQISAKQSLKLGIGDD-----IVFKRYGLV 56
           GL Y+E + V  YE  +N    +  +  L  ++    +  +G   D        ++  L+
Sbjct: 81  GLSYKEKFVVRSYEVGSNKTATVETIANLLQEVGCNHAQSVGFSTDGFATTTTMRKLHLI 140

Query: 57  WVVTDYIIDIERLPKHAEKIVIETEAKAHNKLLCYRYFYIYGE-DGQKIITISSAFVLMD 115
           WV     I+I + P   + + IET  ++  ++   R + +     G+     +S +V+M+
Sbjct: 141 WVTARMHIEIYKYPAWGDVVEIETWCQSEGRIGTRRDWILKDSVTGEVTGRATSKWVMMN 200

Query: 116 FKTRKIHPVLDDITSIY---------------QSQRIKKVIRGPKYHPIGDSKVKQYHVR 160
             TR++  V DD+   Y                ++ +KK+   PK        +     R
Sbjct: 201 QDTRRLQKVSDDVRDEYLVFCPQEPRLAFPEENNRSLKKI---PKLEDPAQYSMIGLKPR 257

Query: 161 YFDLDMNGHVNNSKYLEWMYDVLDLDFLSSHIPKKIDLKYIKEIQYGTDIKSHWYQDGLV 220
             DLDMN HVNN  Y+ W+ + +  + + +H  + I L Y +E Q    +      D L
Sbjct: 258 RADLDMNQHVNNVTYIGWVLESIPQEIVDTHELQVITLDYRRECQQDDVV------DSLT 311

Query: 221 TRHDIIGG 228
           T    IGG
Sbjct: 312 TTTSEIGG 319
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4223> which encodes the amino acid sequence <SEQ ID 4224>. Analysis of this protein sequence reveals the following:

```
     Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.88   Transmembrane   21 -  37 (  21 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.2550(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB71730 GB:U65643 acyl-ACP thioesterase [Myristica fragrans]
 Identities = 41/128 (32%), Positives = 67/128 (52%), Gaps = 11/128 (8%)

Query: 33  FIFMIKRGGLLVDILAYFALLNPDTRKVATIPEDLVAPFETDFVKKLHRV-----PKMPL 87
           F+   K G +L    + + ++N  TR+++ IPE++    E  FV+  H V      K+P
Sbjct: 147 FLRDCKTGEILTRATSVWVMMNKRTRRLSKIPEEVRVEIEPYFVE--HGVLDEDSRKLPK 204

Query: 88  LEQS----IDRDYYVRYFDIDMNGHVNNSKYLDWMYDVLGCEFLKTHQPLKMTLKYVKEV 143
           L  +    I R    R+ D+D+N HVNN KY+ W+ + +    L++H+   MTL+Y KE
Sbjct: 205 LNDNTANYIRRGLAPRWSDLDVNQHVNNVKYIGWILESVPSSLLESHELYGMTLEYRKEC 264

Query: 144 SPGGQITS 151
              G + S
Sbjct: 265 GKDGLLQS 272
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 62/144 (43%), Positives = 94/144 (65%)

Query: 101 GQKIITISSAFVLMDFKTRKIHPVLDDITSIYQSQRIKKVIRGPKYHPIGDSKVKQYHVR 160
           G  ++ I + F L++  TRK+  + +D+ + +++  +KK+ R PK   +  S  + Y+VR
Sbjct: 40  GGLLVDILAYFALLNPDTRKVATIPEDLVAPFETDFVKKLHRVPKMPLLEQSIDRDYYVR 99

Query: 161 YFDLDMNGHVNNSKYLEWMYDVLDLDFLSSHIPKKIDLKYIKEIQYGTDIKSHWYQDGLV 220
           YFD+DMNGHVNNSKYL+WMYDVL  +FL +H P K+ LKY+KE+  G  I S ++ D L
Sbjct: 100 YFDIDMNGHVNNSKYLDWMYDVLGCEFLKTHQPLKMTLKYVKEVSPGGQITSSYHLDQLT 159

Query: 221 TRHDIIGGDAIHAQARIEWQEKKE 244
           + H I     ++AQA IEW+  K+
Sbjct: 160 SYHQITSDGQLNAQAMIEWRAIKQ 183
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1379

A DNA sequence (GBSx1464) was identified in *S.agalactiae* <SEQ ID 4225> which encodes the amino acid sequence <SEQ ID 4226>. This protein is predicted to be coproporphyrinogen III oxidase. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05062 GB:AP001511 coproporphyrinogen III oxidase [Bacillus halodurans]
 Identities = 173/375 (46%), Positives = 248/375 (66%), Gaps = 5/375 (1%)

Query: 5   PTSAYVHIPFCTQICYYCDFSKVFIKNQPVDAYLQALIREFR----SYDITELRTLYIGG 60
           P +AY+HIPFC  ICYYCDF+K ++KNQPV+ YLQAL  E            L+TLY+GG
Sbjct: 2   PKAAYIHIPFCEHICYYCDFNKFYLKNQPVNEYLQALETEMAMVVAEQPTKSLQTLYVGG 61

Query: 61  GTPTSISAVQLDYLLTELSRDLNLNTLEEFTIEANPGDLTVDKIEVLQKSAVNRVSLGVQ 120
           GTPT+++A QL  LL  + R L L+ LEEFT E NP  +  +K++VL+   V+R+S+GVQ
Sbjct: 62  GTPTALTADQLAQLLASIKRTLPLSDLEEFTFEVNPDSIDEEKLDVLRSYGVDRLSIGVQ 121

Query: 121 TFNDKHLKRIGRSHNEAQIYSTIDALKTAGFQNISIDLIYALPGQTMDDVRSNVAKALSL 180
            F    LK IGR+H++  +   ++  + AGF N+S+DL+  LP QT +     + +A +L
Sbjct: 122 AFQPLLLKEIGRTHDQKSVEQAVEKSRQAGFANLSLDLMLGLPKQTPEMFAETLKEAFAL 181
```

```
Query: 181 NIPHLSLYSLILEHHTVFMNKMRRGKLHLPTEDLEAEMFEYIISEMERNGFEHYEISNFT 240
            + HLS YSL +E  TVF N+ R+G+L LP ED E +M+  +  E E++GF+ YEISNF
Sbjct: 182 EVEHLSCYSLKVEAKTVFYNRQRQGRLTLPPEDDEVKMYRQLCYETEKHGFKQYEISNFA 241

Query: 241 KPGFESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRGPIQHYLKGVSEGNARLSE-EVL 299
           K G+ESRHNL+YW+N EYYG GAGA GY+ G+RY N GP+  YL+ + EG   + E   +
Sbjct: 242 KKGYESRHNLVYWNNDEYYGFGAGAHGYVGGVRYMNHGPLPKYLQAMEEGRRPVFESHHV 301

Query: 300 SKNEMMEEELFLGLRKKEGVSIGKFEQKFGTSFEKRYGQIVQELQSDGLLKENNGFIQMT 359
           S+ E MEE++FLGLRK+ GV    F ++FG S    Y + + +L +  LL+  +  +++T
Sbjct: 302 SRVEQMEEQMFLGLRKRSGVEERVFVERFGVSMFSLYEKQIAQLVARCLLERTDDRVRLT 361

Query: 360 KKGLFLGDTVAEKFI 374
            +GL LG+ V E+F+
Sbjct: 362 DEGLLLGNEVFEQFL 376
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4227> which encodes the amino acid sequence <SEQ ID 4228>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3202(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 304/376 (80%), Positives = 343/376 (90%)

Query: 1   MLKKPTSAYVHIPFCTQICYYCDFSKVFIKNQPVDAYLQALIREFRSYDITELRTLYIGG 60
           M KKPTSAYVHIPFCTQICYYCDFSKVFI+NQPVDAYL+ALI+EF SY I +L+TLYIGG
Sbjct: 33  MSKKPTSAYVHIPFCTQICYYCDFSKVFIQNQPVDAYLKALIQEFDSYGIRDLKTLYIGG 92

Query: 61  GTPTSISAVQLDYLLTELSRDLNLNTLEEFTIEANPGDLTVDKIEVLQKSAVNRVSLGVQ 120
           GTPT+I+A QL+YLL  L R+LNL+ LEEFTIEANPGDLT +KI VLQ+SAVNR+SLGVQ
Sbjct: 93  GTPTAITAKQLEYLLNHLERNLNLDDLEEFTIEANPGDLTPEKIAVLQRSAVNRISLGVQ 152

Query: 121 TFNDKHLKRIGRSHNEAQIYSTIDALKTAGFQNISIDLIYALPGQTMDDVRSNVAKALSL 180
           TFN+K LK+IGRSHNE QIYSTI  LKTAGF NISIDLIYALPGQT+D V+ NVAKAL+L
Sbjct: 153 TFNNKQLKQIGRSHNEEQIYSTIANLKTAGFHNISIDLIYALPGQTLDQVKENVAKALAL 212

Query: 181 NIPHLSLYSLILEHHTVFMNKMRRGKLHLPTEDLEAEMFEYIISEMERNGFEHYEISNFT 240
           +IPHLSLYSLILEHHTVFMNKMRRGKL+LPTEDLEAEMFEYIISEME NGFEHYEISNFT
Sbjct: 213 DIPHLSLYSLILEHHTVFMNKMRRGKLNLPTEDLEAEMFEYIISEMEANGFEHYEISNFT 272

Query: 241 KPGFESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRGPIQHYLKGVSEGNARLSEEVLS 300
           KPGFESRHNLMYWDNVEY+G GAGASGYL+GIRY+NR PIQHYLK V  GNARL+EEVL
Sbjct: 273 KPGFESRHNLMYWDNVEYFGCGAGASGYLNGIRYQNRVPIQHYLKAVEAGNARLNEEVLR 332

Query: 301 KNEMMEEELFLGLRKKEGVSIGKFEQKFGTSFEKRYGQIVQELQSDGLLKENNGFIQMTK 360
           K EMMEEELFLGLRKK GVSI +F++KFG SFE+RYG IV+ELQ+ GLL +++ F++MTK
Sbjct: 333 KEEMMEEELFLGLRKKTGVSIQRFQEKFGMSFEERYGNIVRELQNQGLLVKDDAFVRMTK 392

Query: 361 KGLFLGDTVAEKFIVE 376
           KGLFLGD+VAE+FI++
Sbjct: 393 KGLFLGDSVAERFILD 408
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1380

A DNA sequence (GBSx1465) was identified in *S.agalactiae* <SEQ ID 4229> which encodes the amino acid sequence <SEQ ID 4230>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3729(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1381

A DNA sequence (GBSx1466) was identified in *S.agalactiae* <SEQ ID 4231> which encodes the amino acid sequence <SEQ ID 4232>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2989(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4233> which encodes the amino acid sequence <SEQ ID 4234>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2993(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 36/109 (33%), Positives = 58/109 (53%), Gaps = 6/109 (5%)

Query: 9   WAKHKYLVLSKSQKIYLDIRQTLKSPNCT---VLDVQSLIDQAVLLEESPSQVTNAYMHI 65
           WA  KY V++ SQ+ Y  +R+  K     +   VL    LI++A  +  +   + AY H+
Sbjct: 13  WAYQKYWVMAHSQQHYNALRELFKGNQWSEEKVLTFHCLIEEAQAIPPTVKSLRTAYQHV 72

Query: 66  WGYFKNKAERQEKEEFLTLLEKYRKTGYQRRKLLAFLKQLLAKYPNSYL 114
           WGYFK  A ++EK+ F  L  +       +  ++L FL+++ A Y  SYL
Sbjct: 73  WGYFKKVASQEEKDHFKDLDAQLET---KSEEMLCFLQEMTAHYQPSYL 118
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1382

A DNA sequence (GBSx1467) was identified in *S.agalactiae* <SEQ ID 4235> which encodes the amino acid sequence <SEQ ID 4236>. This protein is predicted to be mrsA (mrsA). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane   56 -  72 (  56 -  72)

----- Final Results -----
              bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11970 GB:Z99105 similar to phosphoglucomutase (glycolysis)
           [Bacillus subtilis]
 Identities = 284/451 (62%), Positives = 353/451 (77%), Gaps = 4/451 (0%)

Query: 1   MGKYFGTDGVRGEANVELTPELAFKLGRFGGYVLSQHETDRPRVFVARDTRISGEMLESA 60
           MGKYFGTDGVRG AN ELTPELAFK+GRFGGYVL++ +  RP+V + RDTRISG MLE A
Sbjct: 1   MGKYFGTDGVRGVANSELTPELAFKVGRFGGYVLTK-DKQRPKVLIGRDTRISGHMLEGA 59

Query: 61  LIAGLLSVGIEVYKLGVLATPGVSYLVRTEKASAGVMISASHNPALDNGIKFFGSDGFKL 120
           L+AGLLS+G EV +LGV++TPGVSYL +   A AGVMISASHNP  DNGIKFFG DGFKL
Sbjct: 60  LVAGLLSIGAEVMRLGVISTPGVSYLTKAMDAEAGVMISASHNPVQDNGIKFFGGDGFKL 119

Query: 121 DDDRELEIEALLDAKEDTLPRPSAQGLGTLVDYPEGLRKYEKFMESTGI-DLEGMKVALD 179
            D++E EIE L+D  ED LPRP    LG + DY EG +KY +F++ T   D  G+ VALD
Sbjct: 120 SDEQEAEIERLMDEPEDKLPRPVGADLGLVNDYFEGGQKYLQFLKQTADEDFTGIHVALD 179

Query: 180 TANGAATASARNIFLDLNADISVIGDQPDGLNINDGVGSTHPEQLQSLVRENGSDIGLAF 239
            ANGA ++ A ++F DL+AD+S +G  P+GLNINDGVGSTHPE L + V+E  +D+GLAF
Sbjct: 180 CANGATSSLATHLFADLDADVSTMGTSPNGLNINDGVGSTHPEALSAFVKEKNADLGLAF 239

Query: 240 DGDSDRLIAVDENGEIVDGDKIMFIIGKYLSDKGQLAQNTIVTTVMSNLGFHKALDREGI 299
           DGD DRLIAVDE G IVDGD+IM+I  K+L  +G+L  +T+V+TVMSNLGF+KAL++EGI
Sbjct: 240 DGDGDRLIAVDEKGNIVDGDQIMYICSKHLKSEGRLKDDTVVSTVMSNLGFYKALEKEGI 299

Query: 300 HKAITAVGDRYVVEEMRKSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLTKVMKETGKKL 359
               TAVGDRYVVE M+K GYN+GGEQSGH+I +DYNTTGDG L+AI L   +K TGK L
Sbjct: 300 KSVQTAVGDRYVVEAMKKDGYNVGGEQSGHLIFLDYNTTGDGLLSAIMLMNTLKATGKPL 359

Query: 360 SELASEVTIYPQKLVNIRVENNMKDKAMEVPAIAEIIAKMEEEMDGNGRILVRPSGTEPL 419
           SELA+E+  +PQ LVN+RV +  K K  E    +  +I+++E+EM+G+GRILVRPSGTEPL
Sbjct: 360 SELAAEMQKFPQLLVNVRVTD--KYKVEENEKVKAVISEVEKEMNGDGRILVRPSGTEPL 417

Query: 420 LRVMAEAPTNEAVDYYVDTIADVVRTEIGLD 450
           +RVMAEA T E  D YV+ I +VVR+E+GL+
Sbjct: 418 VRVMAEAKTKELCDEYVNRIVEVVRSEMGLE 448
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4237> which encodes the amino acid sequence <SEQ ID 4238>. Analysis of this protein sequence reveals the following:

```
      Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane   56 -  72 (  56 -  72)

----- Final Results -----
              bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB11970 GB:Z99105 similar to phosphoglucomutase (glycolysis)
          [Bacillus subtilis]
 Identities = 287/451 (63%), Positives = 346/451 (76%), Gaps = 4/451 (0%)

Query: 1   MGKYFGTDGVRGEANVELTPELAFKLGRFGGYVLSQHETERPKVFVARDTRISGEMLESA 60
           MGKYFGTDGVRG AN ELTPELAFK+GRFGGYVL++ + +RPKV + RDTRISG MLE A
Sbjct: 1   MGKYFGTDGVRGVANSELTPELAFKVGRFGGYVLTK-DKQRPKVLIGRDTRISGHMLEGA 59

Query: 61  LIAGLLSVGIEVYKLGVLATPGVSYLVRTEKASAGVMISASHNPALDNGIKFFGNDGFKL 120
           L+AGLLS+G EV +LGV++TPGVSYL +   A AGVMISASHNP  DNGIKFFG DGFKL
Sbjct: 60  LVAGLLSIGAEVMRLGVISTPGVSYLTKAMDAEAGVMISASHNPVQDNGIKFFGGDGFKL 119

Query: 121 ADDQELEIEALLDAPEDTLPRPSAEGLGTLVDYPEGLRKYEKFLVTTGT-DLSGMTVALD 179
           +D+QE EIE L+D PED LPRP    LG + DY EG +KY +FL  T   D +G+ VALD
Sbjct: 120 SDEQEAEIERLMDEPEDKLPRPVGADLGLVNDYFEGGQKYLQFLKQTADEDFTGIHVALD 179

Query: 180 TANGAASVSARDVFLDLNAEIAVIGEKPNGLNINDGVGSTRPEQLQELVKETGADLGLAF 239
            ANGA S  A  +F DL+A+++ +G  PNGLNINDGVGST PE L   VKE  ADLGLAF
Sbjct: 180 CANGATSSLATHLFADLDADVSTMGTSPNGLNINDGVGSTHPEALSAFVKEKNADLGLAF 239

Query: 240 DGDSDRLIAVDETGEIVDGDRIMFIIGKYLSEKGLLAHNTIVTTVMSNLGFHKALDKQGI 299
           DGD DRLIAVDE G IVDGD+IM+I  K+L  +G L  +T+V+TVMSNLGF+KAL+K+GI
Sbjct: 240 DGDGDRLIAVDEKGNIVDGDQIMYICSKHLKSEGRLKDDTVVSTVMSNLGFYKALEKEGI 299

Query: 300 NKAITAVGDRYVVEEMRSSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLAKVMKETGKSL 359
               TAVGDRYVVE M+  GYN+GGEQSGH+I +DYNTTGDG L+AI L   +K TGK L
Sbjct: 300 KSVQTAVGDRYVVEAMKKDGYNVGGEQSGHLIFLDYNTTGDGLLSAIMLMNTLKATGKPL 359

Query: 360 SELAAEVTIYPQKLVNIRVENSMKERAMEVPAIANIIAKMEDEMAGNGRILVRPSGTEPL 419
           SELAAE+  +PQ LVN+RV +  K +  E    +  +I+++E EM G+GRILVRPSGTEPL
Sbjct: 360 SELAAEMQKFPQLLVNVRVTD--KYKVEENEKVKAVISEVEKEMNGDGRILVRPSGTEPL 417

Query: 420 LRVMAEAPTDAEVDYYVDTIADVVRTEIGCD 450
           +RVMAEA T    D YV+ I +VVR+E+G +
Sbjct: 418 VRVMAEAKTKELCDEYVNRIVEVVRSEMGLE 448
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 400/450 (88%), Positives = 429/450 (94%)

Query: 1   MGKYFGTDGVRGEANVELTPELAFKLGRFGGYVLSQHETDRPRVFVARDTRISGEMLESA 60
           MGKYFGTDGVRGEANVELTPELAFKLGRFGGYVLSQHET+RP+VFVARDTRISGEMLESA
Sbjct: 1   MGKYFGTDGVRGEANVELTPELAFKLGRFGGYVLSQHETERPKVFVARDTRISGEMLESA 60

Query: 61  LIAGLLSVGIEVYKLGVLATPGVSYLVRTEKASAGVMISASHNPALDNGIKFFGSDGFKL 120
           LIAGLLSVGIEVYKLGVLATPGVSYLVRTEKASAGVMISASHNPALDNGIKFFG+DGFKL
Sbjct: 61  LIAGLLSVGIEVYKLGVLATPGVSYLVRTEKASAGVMISASHNPALDNGIKFFGNDGFKL 120

Query: 121 DDDRELEIEALLDAKEDTLPRPSAQGLGTLVDYPEGLRKYEKFMESTGIDLEGMKVALDT 180
            DD+ELEIEALLDA EDTLPRPSA+GLGTLVDYPEGLRKYEKF+ +TG DL GM VALDT
Sbjct: 121 ADDQELEIEALLDAPEDTLPRPSAEGLGTLVDYPEGLRKYEKFLVTTGTDLSGMTVALDT 180

Query: 181 ANGAATASARNIFLDLNADISVIGDQPDGLNINDGVGSTHPEQLQSLVRENGSDIGLAFD 240
           ANGAA+ SAR++FLDLNA+I+VIG++P+GLNINDGVGST PEQLQ LV+E G+D+GLAFD
Sbjct: 181 ANGAASVSARDVFLDLNAEIAVIGEKPNGLNINDGVGSTRPEQLQELVKETGADLGLAFD 240

Query: 241 GDSDRLIAVDENGEIVDGDKIMFIIGKYLSDKGQLAQNTIVTTVMSNLGFHKALDREGIH 300
           GDSDRLIAVDE GEIVDGD+IMFIIGKYLS+KG LA NTIVTTVMSNLGFHKALD++GI+
Sbjct: 241 GDSDRLIAVDETGEIVDGDRIMFIIGKYLSEKGLLAHNTIVTTVMSNLGFHKALDKQGIN 300

Query: 301 KAITAVGDRYVVEEMRKSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLTKVMKETGKKLS 360
           KAITAVGDRYVVEEMR SGYNLGGEQSGHVIIMDYNTTGDGQLTAIQL KVMKETGK LS
Sbjct: 301 KAITAVGDRYVVEEMRSSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLAKVMKETGKSLS 360

Query: 361 ELASEVTIYPQKLVNIRVENNMKDKAMEVPAIAEIIAKMEEEMDGNGRILVRPSGTEPLL 420
           ELA+EVTIYPQKLVNIRVEN+MK++AMEVPAIA IIAKME+EM GNGRILVRPSGTEPLL
```

```
Sbjct: 361 ELAAEVTIYPQKLVNIRVENSMKERAMEVPAIANIIAKMEDEMAGNGRILVRPSGTEPLL 420

Query: 421 RVMAEAPTNEAVDYYVDTIADVVRTEIGLD 450
           RVMAEAPT+  VDYYVDTIADVVRTEIG D
Sbjct: 421 RVMAEAPTDAEVDYYVDTIADVVRTEIGCD 450
```

SEQ ID 4236 (GBS402) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 5; MW 78kDa).

GBS402-GST was purified as shown in Figure 218, lane 3-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1383

A DNA sequence (GBSx1468) was identified in *S.agalactiae* <SEQ ID 4239> which encodes the amino acid sequence <SEQ ID 4240>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11969 GB:Z99105 ybbR [Bacillus subtilis]
 Identities = 90/324 (27%), Positives = 167/324 (50%), Gaps = 18/324 (5%)

Query: 1   MKKFFTNKFWLGVVSLFLAILLFLTATATSMNHQDNSKIAG-----ASETYTHTLTDVPI 55
           M KF  N++ + +++L  A+LL++   A + N     K  G      S T   TLTD+P+
Sbjct: 1   MDKFLNNRWAVKIIALLFALLLYV---AVNSNQAPTPKKPGESFFPTSTTDEATLTDIPV 57

Query: 56  DIKYDSDDYFISGYSYGADVYMS-SVNRVKLDSEINEDTRKFKVVADLTNMKPGTHKVPL 114
              YD ++Y ++G     +V +  S + VK   +    T+ F++ AD+ ++K GTHKV L
Sbjct: 58  KAYYDDENYVVTGVPQTVNVTIKGSTSAVKKARQ----TKNFEIYADMEHLKTGTHKVEL 113

Query: 115 KVVNLPSGVNATVSPTTITVTMGKKKTKEFPV-YGHVNDKQIKAGYAVDKMSVDVSKVKV 173
           K  N+  G+  +++P+  TVT+ ++ TK FPV   + N  ++K GY+ ++  V    V++
Sbjct: 114 KAKNVSDGLTISINPSVTTVTIQERTTKSFPVEVEYYNKSKMKKGYSPEQPIVSPKNVQI 173

Query: 174 TSDESIIDRIDHVAANIPDDKVLDDDFNKTVTLQAVTADGTVLASIIHPSKATLSVKVKK 233
           T  +++ID I    A++  +   D+   K   +     DG  L   + PS   ++V V
Sbjct: 174 TGSKNVIDNISLHKASVNLENA-DETIEKEAKVTVYDKDGNALPVDVEPSVIKITVPVTS 232

Query: 234 LTKTVPINLIPVGQFSDSISKINYKLSQEKAVISGTKEALEAISVIN-AEVDISDVTKNT 292
            +K VP  +   G   D +S  N + S  +  + G+++ L+++  I+   +D+S + K++
Sbjct: 233 PSKKVPFKIERTGSLPDGVSIANIESSPSEVTVYGSQDVLDSLEFIDGVSLDLSKINKDS 292

Query: 293 --EKKINLSANNVSVDPAQVTVQL 314
             E  I L      + P++VT+ +
Sbjct: 293 DIEADIPLPDGVKKISPSKVTLHI 316
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4241> which encodes the amino acid sequence <SEQ ID 4242>. Analysis of this protein sequence reveals the following:

```
     Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
```

```
  bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB11969 GB:Z99105 ybbR [Bacillus subtilis]
 Identities = 81/322 (25%), Positives = 154/322 (47%), Gaps = 15/322 (4%)

Query: 1   MKRFLNSRPWLGMVSVFFAILLFLTAASSNH----NNSSSQIYSPIETYTHSLKDVPIDM 56
           M +FLN+R  + ++++ FA+LL++ A +SN            +    T   +L D+P+
Sbjct: 1   MDKFLNNRWAVKIIALLFALLLYV-AVNSNQAPTPKKPGESFFPTSTTDEATLTDIPVKA 59

Query: 57  KYDSDKYFISGYSYGAEVYLT-STNRIKLDSEVNNDTRNFKIVADLTHSHPGTVSVNLRV 115
            YD + Y ++G      V +  ST+ +K   +    T+NF+I AD+ H   GT  V L+
Sbjct: 60  YYDDENYVVTGVPQTVNVTIKGSTSAVKKARQ----TKNFEIYADMEHLKTGTHKVELKA 115

Query: 116 ENLPSGVTATVSPDKISVTIGKKESKVFPVRGS-VDAKQIANGYEISKIETGVNKVEVTS 174
           +N+  G+T +++P   +VTI ++ +K FPV     +  ++  GY   +       V++T
Sbjct: 116 KNVSDGLTISINPSVTTVTIQERTTKSFPVEVEYYNKSKMKKGYSPEQPIVSPKNVQITG 175

Query: 175 DESTIALIDHVVAKLPDDQVLDRNYSSRVTLQAVSADGTILASAIDPAKTNLSVAVKKIT 234
            ++ I  I    A +  +   D        +     DG  L   ++P+   ++V V   +
Sbjct: 176 SKNVIDNISLHKASVNLENA-DETIEKEAKVTVYDKDGNALPVDVEPSVIKITVPVTSPS 234

Query: 235 KSVPIRVEAVGMMDDSLSDIQYKLSKQTAVISGSREVLEDIDEII-AEVNISDVTKNT-- 291
           K VP ++E  G + D +S    + S     + GS++VL+ ++ I    +++S + K++
Sbjct: 235 KKVPFKIERTGSLPDGVSIANIESSPSEVTVYGSQDVLDSLEFIDGVSLDLSKINKDSDI 294

Query: 292 SKTVSLSSSQVSIEPSVVTVQL 313
              + L      I PS VT+ +
Sbjct: 295 EADIPLPDGVKKISPSKVTLHI 316
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 198/319 (62%), Positives = 251/319 (78%), Gaps = 1/319 (0%)

Query: 1   MKKFFTNKFWLGVVSLFLAILLFLTATATSMNHQDNSKIAGASETYTHTLTDVPIDIKYD 60
           MK+F  ++ WLG+VS+F AILLFLTA A+S ++  +S+I    ETYTH+L DVPID+KYD
Sbjct: 1   MKRFLNSRPWLGMVSVFFAILLFLTA-ASSNHNNSSSQIYSPIETYTHSLKDVPIDMKYD 59

Query: 61  SDDYFISGYSYGADVYMSSVNRVKLDSEINEDTRKFKVVADLTNMKPGTHKVPLKVVNLP 120
           SD YFISGYSYGA+VY++S NR+KLDSE+N DTR FK+VADLT+  PGT  V L+V NLP
Sbjct: 60  SDKYFISGYSYGAEVYLTSTNRIKLDSEVNNDTRNFKIVADLTHSHPGTVSVNLRVENLP 119

Query: 121 SGVNATVSPTTITVTMGKKKTKEFPVYGHVNDKQIKAGYAVDKMSVDVSKVKVTSDESII 180
           SGV ATVSP  I+VT+GKK++K FPV G V+ KQI  GY + K+   V+KV+VTSDES I
Sbjct: 120 SGVTATVSPDKISVTIGKKESKVFPVRGSVDAKQIANGYEISKIETGVNKVEVTSDESTI 179

Query: 181 DRIDHVAANIPDDKVLDDDFNKTVTLQAVTADGTVLASIIHPSKATLSVKVKKLTKTVPI 240
             IDHV A +PDD+VLD +++  VTLQAV+ADGT+LAS I P+K  LSV VKK+TK+VPI
Sbjct: 180 ALIDHVVAKLPDDQVLDRNYSSRVTLQAVSADGTILASAIDPAKTNLSVAVKKITKSVPI 239

Query: 241 NLIPVGQFSDSISKINYKLSQEKAVISGTKEALEAISVINAEVDISDVTKNTEKKINLSA 300
            +  VG   DS+S I YKLS++ AVISG++E LE I  I AEV+ISDVTKNT K ++LS+
Sbjct: 240 RVEAVGMMDDSLSDIQYKLSKQTAVISGSREVLEDIDEIIAEVNISDVTKNTSKTVSLSS 299

Query: 301 NNVSVDPAQVTVQLTTTKK 319
           + VS++P+ VTVQLTTTKK
Sbjct: 300 SQVSIEPSVVTVQLTTTKK 318
```

SEQ ID 4240 (GBS99) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 6; MW 35.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 9; MW 60.7kDa).

The GBS99-GST fusion product was purified (Figure 197, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 293), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1384

A DNA sequence (GBSx1469) was identified in *S.agalactiae* <SEQ ID 4243> which encodes the amino acid sequence <SEQ ID 4244>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0503(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1385

A DNA sequence (GBSx1470) was identified in *S.agalactiae* <SEQ ID 4245> which encodes the amino acid sequence <SEQ ID 4246>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.50   Transmembrane   20 -  36 (  18 -  46)
   INTEGRAL    Likelihood = -7.64   Transmembrane   48 -  64 (  42 -  68)
   INTEGRAL    Likelihood = -3.40   Transmembrane   80 -  96 (  80 -  96)

----- Final Results -----
               bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11968 GB:Z99105 alternate gene name: ybbQ~similar to
           hypothetical proteins [Bacillus subtilis]
 Identities = 125/253 (49%), Positives = 186/253 (73%), Gaps = 5/253 (1%)

Query: 27  MDIIIVAVLIYKFIKALAGTKIMSLIQGVILFIIIRFVSEWIGLTTITFLMNQIVTYGVI 86
            +DI++V  +IYK I  + GTK + L++G+++ +++R  S+++GL+T+ +LM+Q +T+G +
Sbjct: 16  VDILLVWYVIYKLIMVIRGTKAVQLLKGIVVIVLVRMASQYLGLSTLQWLMDQAITWGFL 75

Query: 87  AGVVIFAPEIRTGLEKFGRTPQLFTQRSQLSSDE---KLVDALVKAVAYMSPRKIGALIS 143
            A ++IF PE+R  LE+ GR    F  RS    +E   K ++A+ KA+ YM+ R+IGAL++
Sbjct: 76  AIIIIFQPELRRALEQLGRGR--FFSRSGTPVEEAQQKTIEAITKAINYMAKRRIGALLT 133

Query: 144 IERTQTLQEYIATGIPLDADISSELLINIFIPNTPLHDGAVIVKDKKIATACSYLPLSES 203
            IER   + +YI TGIPL+A +SSELLINIFIPNTPLHDGAVI+K+ +IA A  YLPLSES
Sbjct: 134 IERDTGMGDYIETGIPLNAKVSSELLINIFIPNTPLHDGAVIMKNNEIAAAACYLPLSES 193

Query: 204 SSISKEFGTRHRAAIGLSENSDALTVIVSEETGGISVALKGEFLHDLSKDSFEAILRTQL 263
```

```
             ISKE GTRHRAA+G+SE +D+LT+IVSEETGG+SVA  G+   +L++++ + +L  +
Sbjct: 194 PFISKELGTRHRAAVGISEVTDSLTIIVSEETGGVSVAKNGDLHRELTEEALKEMLEAEF 253

Query: 264 IQNQEENSKLAWY 276
             +N  + S   WY
Sbjct: 254 KKNTRDTSSNRWY 266
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4247> which encodes the amino acid sequence <SEQ ID 4248>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.64   Transmembrane   20 -  36 (  19 -  40)
   INTEGRAL    Likelihood = -6.21   Transmembrane   48 -  64 (  47 -  68)
   INTEGRAL    Likelihood = -2.07   Transmembrane   76 -  92 (  76 -  92)

----- Final Results -----
                 bacterial membrane --- Certainty=0.3654(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB03984 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 117/255 (45%), Positives = 178/255 (68%), Gaps = 6/255 (2%)

Query: 19  PWL-LAVHLLDILIVAYLIYRFIKALTGTKIMSLVQGVIFFLVLRFIAEWIGFTTITYLM 77
            PWL     +LDIL+V Y+IY+ I  + GT+ + L++G+   L++  I+ +    T+ +++
Sbjct: 8   PWLNYLTQILDILVVTYVIYKAIMIIRGTRAVQLLKGITVILIVYAISIFFNLRTLGWIV 67

Query: 78  NQVITYGVIAGVVIFTPEIRAGLEKFGRSTQVFLQKQYVSSESAL---VDALIKSVAYMG 134
            NQ ITYG++A ++IF PE+R  LE+ GR    F   +  + E  +   +DA++K+  YMG
Sbjct: 68  NQAITYGLLAVIIIFQPELRRALEQLGRGR--FFASRTANEEETMKKTIDAIVKASTYMG 125

Query: 135 PRKIGALIAIEQTQTLQEYIATGIPLNADISSQLLINIFIPNTPLHDGAVIVGQNKIVAA 194
             R+IGALI++E+   + +Y+ TGIP+NA+++S+LLIN FIPNTPLHDGAVI+  + I+AA
Sbjct: 126 KRRIGALISMERETGMTDYVETGIPMNANLTSELLINTFIPNTPLHDGAVIINNDTILAA 185

Query: 195 CAYLPLSESKAISKEFGTRHRAAIGLSENSDALTIIVSEETGAISVTRKGQFLHDLSTDE 254
              YLPLSE+  ISKE GTRHRAA+G+SE +D LTI+VSEETG IS+T+ G+   DL  ++
Sbjct: 186 ACYLPLSENPFISKELGTRHRAALGVSEVTDCLTIVVSEETGHISLTKNGELHRDLDEEQ 245

Query: 255 FETVLRTYLMSNSNV 269
              ++L   L+S + +
Sbjct: 246 LRSLLEAELISEAKM 260
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/283 (71%), Positives = 239/283 (84%), Gaps = 2/283 (0%)

Query: 1   MDIFSAIDSKFWASIMENPWMILIHLMDIIIVAVLIYKFIKALAGTKIMSLIQGVILFII 60
            M+  S+ID KF  S+  +PW++ +HL+DI+IVA LIY+FIKAL GTKIMSL+QGVI F++
Sbjct: 1   MNNLSSIDIKFLLSLFADPWLLAVHLLDILIVAYLIYRFIKALTGTKIMSLVQGVIFFLV 60

Query: 61  IRFVSEWIGLTTITFLMNQIVTYGVIAGVVIFAPEIRTGLEKFGRTPQLFTQRSQLSSDE 120
            +RF++EWIG TTIT+LMNQ++TYGVIAGVVIF PEIR GLEKFGR+ Q+F Q+  +SS+
Sbjct: 61  LRFIAEWIGFTTITYLMNQVITYGVIAGVVIFTPEIRAGLEKFGRSTQVFLQKQYVSSES 120

Query: 121 KLVDALVKAVAYMSPRKIGALISIERTQTLQEYIATGIPLDADISSELLINIFIPNTPLH 180
             LVDAL+K+VAYM PRKIGALI+IE+TQTLQEYIATGIPL+ADISS+LLINIFIPNTPLH
Sbjct: 121 ALVDALIKSVAYMGPRKIGALIAIEQTQTLQEYIATGIPLNADISSQLLINIFIPNTPLH 180

Query: 181 DGAVIVKDKKIATACSYLPLSESSSISKEFGTRHRAAIGLSENSDALTVIVSEETGGISV 240
            DGAVIV   KI  AC+YLPLSES +ISKEFGTRHRAAIGLSENSDALT+IVSEETG ISV
Sbjct: 181 DGAVIVGQNKIVAACAYLPLSESKAISKEFGTRHRAAIGLSENSDALTIIVSEETGAISV 240

Query: 241 ALKGEFLHDLSKDSFEAILRTQLIQNQEENSKLAWYNQLLRRK 283
```

```
             KG+FLHDLS D FE +LRT L+ N   N  L WY ++L  K
Sbjct: 241 TRKGQFLHDLSTDEFETVLRTYLMSN--SNVTLPWYKKILGGK 281
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1386

A DNA sequence (GBSx1471) was identified in *S.agalactiae* <SEQ ID 4249> which encodes the amino acid sequence <SEQ ID 4250>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane    33 -  49 (  33 -  49)

----- Final Results -----
               bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1387

A DNA sequence (GBSx1472) was identified in *S.agalactiae* <SEQ ID 4251> which encodes the amino acid sequence <SEQ ID 4252>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1001(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9781> which encodes amino acid sequence <SEQ ID 9782> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC84012 GB:AF080002 UDP-N-acetylmuramyl tripeptide synthetase
           MurC [Heliobacillus mobilis]
 Identities = 143/442 (32%), Positives = 229/442 (51%), Gaps = 17/442 (3%)

Query: 12  GKSAHYLLSKMGRGST-YPGSLALKFDKDILDTIAKDYE--IVVVTGTNGKTLTTALTVG 68
            GK+A +L  + G G T +PG +  +    IL  +A+  +   +VVTGTNGKT T+ +
Sbjct: 2   GKTAIWLNRRFGHGGTSFPGGIGRRVAPQILTALARQLKRGAMVVTGTNGKTTTSKMLAA 61

Query: 69  ILKEAFGQVVTNPSGANMITGIVSTFLTAKKSKSG--KKIAVLEIDEASLPRITQYIKPS 126
            I++++   +  N +GAN++ GI + F+ +          + ++E+DEA++P++ + ++P
Sbjct: 62  IVEKSSLTLTHNRAGANLVGGITTAFIDSATIGGSITSDLGIIEVDEATIPQLVREVQPK 121

Query: 127 LFVFTNIFRDQMDRYGEIYTTYQMILDGAANAP-QATILANGDSPLFNS--KSVTNPVQF 183
              V TN FRDQ+DR+GE+  T  ++ +     P Q+  + N D PL  S  K     V +
Sbjct: 122 GVVVTNFFRDQLDRFGELDKTVSLVGEALRLLPVQSIAVLNADDPLVASLGKDFPGRVLY 181

Query: 184 YGFNTDKHEPRLAHYNTEGILCPKCQAILTYRLNTYANLGDYTCPNCDFERPNLDYALTR 243
            +G +   +  R    + E   C  C   LTY    +  LG Y C +C FERP     +T
```

```
Sbjct: 182 FGIDDRSYGAREMLQSAETRFCRLCGHPLTYDWFFFGQLGHYRCSHCGFERPEPKIKVTG 241

Query: 244 LTHLTNTSSGFVIDGQ----QYNINVGGLYNIYNALAAVSVAEYFGVEPSQIKDGFDKSR 299
           +       S F ++      Q  ++  G YNIYNALAA++ A    +    I+ G    R
Sbjct: 242 IQLKGEEGSAFTVETPRGTWQLELSTPGFYNIYNALAAIASAIRLDLPEKAIRAGLQGYR 301

Query: 300 AVFGRQETFTIGN-KKCTLVLIKNPVGASQALDMIKLAPYPFSLSVLLNANYADGIDTSW 358
             FGR E   + + ++  L LIKNP G  + +  +     P  L V++N N ADG D SW
Sbjct: 302 TNFGRMERIELEDGRRAFLALIKNPTGCDEVIRTLVQNRGPKRLLVIINDNAADGRDISW 361

Query: 359 IWDANFETI--LTMNIPEIFAGGVRHSEIARRLRVTGYDEKRIK-QADKLQDIMTMIEQQ 415
           +WDA+FE++  +   +  +F  G+R  ++A RL  TG   + I+ +A+    I + +E
Sbjct: 362 LWDADFESLEPVYPELRSVFTSGLRGEDMALRLNYTGIPAESIRYEANVESAIRSALEMT 421

Query: 416 ET-EHAYILATYTAMLEFREIL 436
           E  E  YIL TYTA+LE +  L
Sbjct: 422 EPGETLYILPTYTALLESKAAL 443
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4253> which encodes the amino acid sequence <SEQ ID 4254>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 343/446 (76%), Positives = 393/446 (87%)

Query: 1   MKINTALGVAAGKSAHYLLSKMGRGSTYPGSLALKFDKDILDTIAKDYEIVVVTGTNGKT 60
           MK+ T LG+ AGK+A  +L+K+GRGSTYPG LAL  DKDIL  ++KDY+IVVVTGTNGKT
Sbjct: 1   MKMKTLLGIIAGKAAQSILTKLGRGSTYPGRLALACDKDILKDLSKDYDIVVVTGTNGKT 60

Query: 61  LTTALTVGILKEAFGQVVTNPSGANMITGIVSTFLTAKKSKSGKKIAVLEIDEASLPRIT 120
           LTTALTVGILKEAFG+++TNPSGANMITGI STFL AKK KS ++IAVLEIDEASLPRIT
Sbjct: 61  LTTALTVGILKEAFGEIITNPSGANMITGITSTFLAAKKGKSERQIAVLEIDEASLPRIT 120

Query: 121 QYIKPSLFVFTNIFRDQMDRYGEIYTTYQMILDGAANAPQATILANGDSPLFNSKSVTNP 180
            Y+KPSLFV+TNIFRDQMDRYGEIYTTYQMI+DGA NAP+ATILANGDSP+F+SK + NP
Sbjct: 121 TYLKPSLFVYTNIFRDQMDRYGEIYTTYQMIVDGARNAPKATILANGDSPIFSSKDIVNP 180

Query: 181 VQFYGFNTDKHEPRLAHYNTEGILCPKCQAILTYRLNTYANLGDYTCPNCDFERPNLDYA 240
           VQ+YGF+T KH P+LAHYNTEGILCPKC+ IL YRLNTYANLGD+ C NC F+RP LDY
Sbjct: 181 VQYYGFDTAKHAPQLAHYNTEGILCPKCEHILQYRLNTYANLGDFVCLNCQFQRPTLDYQ 240

Query: 241 LTRLTHLTNTSSGFVIDGQQYNINVGGLYNIYNALAAVSVAEYFGVEPSQIKDGFDKSRA 300
           LT LT +T+ SS FVIDGQ Y INVGGLYNIYNALAAVSVAE+FGV P +IK GF+KS+A
Sbjct: 241 LTELTAITHQSSEFVIDGQNYKINVGGLYNIYNALAAVSVAEFFGVSPEKIKAGFNKSKA 300

Query: 301 VFGRQETFTIGNKKCTLVLIKNPVGASQALDMIKLAPYPFSLSVLLNANYADGIDTSWIW 360
           VFGRQETFT+G+K CTL+LIKNPVGASQAL+MI+LA YPFSLSVLLNANYADGIDTSWIW
Sbjct: 301 VFGRQETFTVGDKSCTLILIKNPVGASQALEMIQLADYPFSLSVLLNANYADGIDTSWIW 360

Query: 361 DANFETILTMNIPEIFAGGVRHSEIARRLRVTGYDEKRIKQADKLQDIMTMIEQQETEHA 420
           DANFE I  M I EI AGGVRHSEIARRLRVTG+D+ +IKQA+KL+ I+  IE+QE +HA
Sbjct: 361 DANFELITQMPITEINAGGVRHSEIARRLRVTGFDDTKIKQAEKLEQIIETIEKQEAKHA 420

Query: 421 YILATYTAMLEFREILANHNAIRKEM 446
           YILATYTAMLEFR +LA+ + + KEM
Sbjct: 421 YILATYTAMLEFRSLLADRHVVEKEM 446
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1388

A DNA sequence (GBSx1473) was identified in *S.agalactiae* <SEQ ID 4255> which encodes the amino acid sequence <SEQ ID 4256>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3010(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC84011 GB:AF080002 cobyric acid synthase CobQ [Heliobacillus
           mobilis]
 Identities = 89/250 (35%), Positives = 129/250 (51%), Gaps = 9/250 (3%)

Query: 11  TKDYKYTLNVAHLYGNLLNTYGDNGNILMMKYVGEKLGCQMTFDIVSLEDRFDPNYYQMA 70
            +K    TL + HLY +LLN YGD GNI+ ++   E  G  +     SL ++   +   +
Sbjct: 2   SKTSNRTLTLIHLYPDLLNLYGDRGNIITLRRRCEWRGITLQVHSASLGEKAAFDDADLV 61

Query: 71  FFGGGQDYEQAIVARDLPSKKEDINKFIQNNGV-VLAICGGFQLLGQYYIQANGERIEGI 129
            F GGG D EQ ++ +D    K          G+ +L++CGG+QLLG YY    GE + G+
Sbjct: 62  FMGGGSDREQTLLFQDFQQHKGPALVEAAEGGLPLLSVCGGYQLLGLYYRTHTGEEMPGL 121

Query: 130 GVMGHYTLNQNNNRYIGDIKIHNDEFNE--TYYGFENHQGRTFLSEDE--KPLGTVIYGN 185
            G+   +T    + R IG++        E  T  GFENH GRTFL      +PL  V  G
Sbjct: 122 GLFDAWT-EAGSTRLIGNVVAQAPLLGEQATLVGFENHSGRTFLGSRGGIQPLAQVTAGF 180

Query: 186 GNNKEDGTEGVHYKNVFGSYFHGPILSRNANLAYRLVATALRNKYG---KEIVLPSYEEI 242
            GNN +D  EG  YKN  G+Y HGP+L +N  LA  L++ AL  +YG      +  ++E
Sbjct: 181 GNNGDDQGEGAVYKNAVGTYLHGPVLPKNPALADWLLSKALERRYGGGSLSTLQDTWENR 240

Query: 243 LSLEIPEEYG 252
              L + + +G
Sbjct: 241 AHLSVAQRFG 250
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4257> which encodes the amino acid sequence <SEQ ID 4258>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2586(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 197/260 (75%), Positives = 224/260 (85%)

Query: 1   MTYTSLKSPTTKDYKYTLNVAHLYGNLLNTYGDNGNILMMKYVGEKLGCQMTFDIVSLED 60
            MTYTSLKSP  +DY Y L +AHLYGNL+NTYGDNGNILM+KYV EKLG ++T DIVS+ D
Sbjct: 1   MTYTSLKSPENQDYIYDLTIAHLYGNLMNTYGDNGNILMLKYVAEKLGARVTVDIVSIND 60

Query: 61  RFDPNYYQMAFFGGGQDYEQAIVARDLPSKKEDINKFIQNNGVVLAICGGFQLLGQYYIQ 120
             F+ + Y + FFGGGQDYEQ+IVA+DLPSKK  +  +I NN VVLAICGGFQLLGQYY+Q
Sbjct: 61  TFEQDDYDIVFFGGGQDYEQSIVAKDLPSKKAALADYIANNKVVLAICGGFQLLGQYYVQ 120
```

```
Query: 121 ANGERIEGIGVMGHYTLNQNNNRYIGDIKIHNDEFNETYYGFENHQGRTFLSEDEKPLGT 180
           ANG +I+G+G+MGHYTLNQ+ NR+IGDIKIHNDEFNETYYGFENHQGRTFLS DEKPLG
Sbjct: 121 ANGVKIDGLGIMGHYTLNQHQNRFIGDIKIHNDEFNETYYGFENHQGRTFLSGDEKPLGR 180

Query: 181 VIYGNGNNKEDGTEGVHYKNVFGSYFHGPILSRNANLAYRLVATALRNKYGKEIVLPSYE 240
           V+YGNGNNKED TEGVHYKNV+GSYFHGPILSRN NLAYRLV TAL+ KYG  I LPSY+
Sbjct: 181 VVYGNGNNKEDQTEGVHYKNVYGSYFHGPILSRNVNLAYRLVTTALKKKYGSAISLPSYD 240

Query: 241 EILSLEIPEEYGDVKSKADF 260
           +IL  EI EEY D+KSKA F
Sbjct: 241 DILKQEITEEYADLKSKASF 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1389

A DNA sequence (GBSx1474) was identified in *S.agalactiae* <SEQ ID 4259> which encodes the amino acid sequence <SEQ ID 4260>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1701(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04402 GB:AP001509 lipoate-protein ligase [Bacillus halodurans]
 Identities = 153/316 (48%), Positives = 212/316 (66%), Gaps = 3/316 (0%)

Query: 10  DPAYNVALEAYAFQKLTDIDEIFIL-WINEPAIIIGRHQNTIQEINKEFIDKNGIHVVRR 68
           DP  N+A+E YA + L DI+E ++L +INEP+IIIGR+QNTI+EIN E+++ NGIHVVRR
Sbjct: 11  DPRINLAIEEYALKNL-DINETYLLFYINEPSIIIGRNQNTIEEINTEYVESNGIHVVRR 69

Query: 69  LSGGGAVYHDLNNLNYTIISNNTQEGAFDFQTFSKPVIDTLAKLGVKAEFTGRNDL-EIN 127
           LSGGGAVYHD  NLN++ I+ +  E   +FQ F+ PVI  LAKLGV AE  GRND+   +
Sbjct: 70  LSGGGAVYHDHGNLNFSFITKDDGESFSNFQKFTDPVIKALAKLGVTAELKGRNDIIASD 129

Query: 128 GQKFAGNAQAYYKGRMMHHGCLLFDVDMSVLGQALKVSKDKIESKGIKSVRARVTNIVDH 187
           G+K +GNAQ   KGRM  HG LLFD ++   +  AL VSKDKIESKGIKS+R+RV NI  +
Sbjct: 130 GRKISGNAQFSTKGRMFSHGTLLFDSEIDHVVSALNVSKDKIESKGIKSIRSRVANISEF 189

Query: 188 LSDKITVQEFSDAILAQMKEEYPEMDEYVLSDAELSEIQAMRDNQFATWDWTYGKAPEYT 247
           L++KI++ +F   +L  + +    + EY L+  + +EI  +   ++  WDW YGK+P +
Sbjct: 190 LTEKISIDQFRSLLLESIFDGQANIQEYKLTADDWAEIHELSKERYQNWDWNYGKSPAFN 249

Query: 248 IERGVRYPAGKITTYANVENSTIKSVKIFGDFFGVKPVDDIEKMLEGVRYDYKDVLAALK 307
           ++   R+P G I     V+  TI+  KIFGDFFG   V D+E  L G+RY+  D+  AL
Sbjct: 250 LQHSHRFPVGNIDIRLEVKGGTIQQCKIFGDFFGTGDVRDLEDRLVGIRYERADIEQALA 309

Query: 308 TVDTSQYFSRMTPEEI 323
            VD   YF ++  ++I
Sbjct: 310 DVDVKTYFGQVEKDDI 325
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4261> which encodes the amino acid sequence <SEQ ID 4262>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1271(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 249/328 (75%), Positives = 292/328 (88%)

Query: 1   MKYIVNTSNDPAYNVALEAYAFQKLTDIDEIFILWINEPAIIIGRHQNTIQEINKEFIDK 60
           MKYIVN S++PA+N+ALEAYAF++L + DE+FILWINEPAIIIG+HQNTIQEINKE+ID+
Sbjct: 1   MKYIVNKSHNPAFNIALEAYAFRELVEEDELFILWINEPAIIIGKHQNTIQEINKEYIDE 60

Query: 61  NGIHVVRRLSGGGAVYHDLNNLNYTIISNNTQEGAFDFQTFSKPVIDTLAKLGVKAEFTG 120
           +GIHVVRRLSGGGAVYHDLNNLNYTIISN T EGAFDF+TFS+PVI TLA LGV A FTG
Sbjct: 61  HGIHVVRRLSGGGAVYHDLNNLNYTIISNKTAEGAFDFKTFSQPVIATLADLGVTANFTG 120

Query: 121 RNDLEINGQKFAGNAQAYYKGRMMHHGCLLFDVDMSVLGQALKVSKDKIESKGIKSVRAR 180
           RND+EI+G+K  GNAQAYYKGRMMHHGCLLFDVDM+VLG ALKVSKDKIESKG+KSVRAR
Sbjct: 121 RNDIEIDGKKICGNAQAYYKGRMMHHGCLLFDVDMTVLGDALKVSKDKIESKGVKSVRAR 180

Query: 181 VTNIVDHLSDKITVQEFSDAILAQMKEEYPEMDEYVLSDAELSEIQAMRDNQFATWDWTY 240
           VTNI++ L +KITV+EFSD ILA+MKE YP+M EYVLS+ EL++I+     QF +WDWTY
Sbjct: 181 VTNILNELPEKITVEEFSDKILAKMKETYPDMTEYVLSEDELAKIEQSAKEQFGSWDWTY 240

Query: 241 GKAPEYTIERGVRYPAGKITTYANVENSTIKSVKIFGDFFGVKPVDDIEKMLEGVRYDYK 300
           GKAPEYTIER VRYPAGKI+T+ANVENS IK++KI+GDFFG+K V DIE +L G +Y+Y+
Sbjct: 241 GKAPEYTIERNVRYPAGKISTFANVENSIIKNLKIYGDFFGIKDVQDIENLLIGCKYEYR 300

Query: 301 DVLAALKTVDTSQYFSRMTPEEITKAIV 328
           DV   LKT+DT+QYFSRMT EE+ KAIV
Sbjct: 301 DVFERLKTIDTTQYFSRMTVEEVAKAIV 328
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1390

A DNA sequence (GBSx1475) was identified in *S.agalactiae* <SEQ ID 4263> which encodes the amino acid sequence <SEQ ID 4264>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane  294 - 310 ( 294 - 312)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA21748 GB:L31844 dihydrolipoamide dehydrogenase [Clostridium
           magnum]
 Identities = 229/589 (38%), Positives = 339/589 (56%), Gaps = 25/589 (4%)

Query: 1   MAFDVIMPKLGVDMQEGEILEWKKNEGDTVNEGDVLLEIMSDKTNMEIEAEDTGVLLKIV 60
           MA  V+MPKLG+ M EG ++ WKK EGD V  G++L E+ +DK   E+E+ D G++ K++
Sbjct: 1   MAKIVVMPKLGLTMTEGTLVTWKKAEGDQVKVGEILFEVSTDKLTNEVESSDEGIVRKLL 60

Query: 61  HQAGDVVPVTEVIAYIGEEGEEVGTSSPSADATITAEDGQSVSGPAAPSQETVAAATPKE 120
              GDVV     +A IG   E++ +            +G S    +A   +T A   PK+
Sbjct: 61  VNEGDVVECLNPVAIIGSADEDISSLL----------NGSSEGSGSAEQSDTKA---PKK 107

Query: 121 ELAADEY--DIVVVGGGPAGYYAAIRGAQLGGKIAIVEKTEFGGTCLNVGCIPTKTYLKN 178
           E+ A +   ++VV+GGGP GY AAIR AQLG K+ ++EK   GGTCLNVGCIPTK  L +
Sbjct: 108 EVEAVKGGDNLVVIGGGPGGYVAAIRAAQLGAKVTLIEKESLGGTCLNVGCIPTKVLLHS 167
```

```
Query: 179 AEILDGLKVAAGRGINLASTNYAIDMDKTVAFKNSVVKTLTGGVRGLLKANKVEIFNGLG 238
           +++L  +K     GI++  +   ++       K  V+K L  GV GLL  NKV++  G
Sbjct: 168 SQLLTEMKEGDKLGIDIEGS-IVVNWKHIQKRKKIVIKKLVSGVSGLLTCNKVKVIKGTA 226

Query: 239 QVNPDKSVVIGDK-----VIKGRNVVLATGSKVSRINIPGIESPLVLTSDDILDLREIPK 293
           +     ++++  +      +   N ++ATGS      I G +   V+ S   L L   P+
Sbjct: 227 KFESKDTILVTKEDGVAEKVNFDNAIIATGSMPFIPEIEGNKLSGVIDSTGALSLESNPE 286

Query: 294 SLAVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILAKKGMKIKTS 353
           S+A++GGGV+G+E   ++ S G  V++IEM   I+P MD+E+S   +  L + G+ I  +
Sbjct: 287 SIAIIGGGVIGVEFASIFNSLGCKVSIIEMLPHILPPMDREISEIAKAKLIRDGININNN 346

Query: 354 VGVSEIVEANNQLTLKL--NNGEEVV-ADKALLSIGRVPQMNGLENLEPELEMERGRIKV 410
             V+ I +  + L +    + GEE +  +K L+++GR  + GL+  +  ++ E G I V
Sbjct: 347 CKVTRIEQGEDGLKVSFIGDKGEESIDVEKVLIAVGRRSNIEGLDVEKIGVKTEGGSIIV 406

Query: 411 NAYQETSIPGIYAPGDVNGTRMLAHAAYRMGEVAAENALGGNKRKAHLDFTPAAVYTHPE 470
           N   ET++ GIYA GD  G  MLAH A   G VAAEN +G NK K      PA VYT PE
Sbjct: 407 NDKMETNVEGIYAIGDCTGKIMLAHVASDQGVVAAENIMGQNK-KMDYKTVPACVYTKPE 465

Query: 471 VAMVGMTEEQAREQYGDILVGKNSFTGNGRAIASNEAHGFVKVIAEPKYKEILGVHIIGP 530
           +A VG+TEEQA+E+  D  VGK     NG+++  NE  G +K+I + KY+EILGVHI+GP
Sbjct: 466 LASVGLTEEQAKEKGIDYKVGKFQLAANGKSLIMNETGGVIKIITDKKYEEILGVHILGP 525

Query: 531 AAAELINEASTIMENELTVYDVAQSIHGHPTFSEVMYEAFLDVLGEAIH 579
            A +LI EA+  +  E T+ ++  ++H HPT  E M EA L V  +AIH
Sbjct: 526 RATDLITEAALALRLEATLEEIITTVHAHPTVGEAMKEAALAVNNQAIH 574
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1819> which encodes the amino acid sequence <SEQ ID 1820>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane  297 - 313 ( 297 - 315)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 497/591 (84%), Positives = 538/591 (90%), Gaps = 10/591 (1%)

Query: 1   MAFDVIMPKLGVDMQEGEILEWKKNEGDTVNEGDVLLEIMSDKTNMEIEAEDTGVLLKIV 60
           MA ++IMPKLGVDMQEGEI+EWKK EGDTVNEGD+LLEIMSDKTNME+EAED+GVLLKI
Sbjct: 1   MAVEIIMPKLGVDMQEGEIIEWKKQEGDTVNEGDILLEIMSDKTNMELEAEDSGVLLKIT 60

Query: 61  HQAGDVVPVTEVIAYIGEEGEEVGTSSPSA---DATITAEDGQS--VSGPAAPSQETVAA 115
            QAG+ VPVTEVI YIG EGE V  SSP+A   +   T ED ++  +  P AP+Q   A+
Sbjct: 61  RQAGETVPVTEVIGYIGAEGESVEVSSPAASDVNVARTTEDLEAAGLEVPKAPAQ--AAS 118

Query: 116 ATPKEELAADEYDIVVVGGGPAGYYAAIRGAQLGGKIAIVEKTEFGGTCLNVGCIPTKTY 175
           A PK  LA DEYDI+VVGGGPAGYYAAIRGAQLGGKIAIVEK+EFGGTCLNVGCIPTKTY
Sbjct: 119 AAPKAALADDEYDIIVVGGGPAGYYAAIRGAQLGGKIAIVEKSEFGGTCLNVGCIPTKTY 178

Query: 176 LKNAEILDGLKVAAGRGINLASTNYAIDMDKTVAFKNSVVKTLTGGVRGLLKANKVEIFN 235
           LKNAEILDG+K+AAGRGINLASTNY IDMDKTV FKN+VVKTLTGGV+GLLKANKV IFN
Sbjct: 179 LKNAEILDGIKIAAGRGINLASTNYTIDMDKTVDFKNTVVKTLTGGVQGLLKANKVTIFN 238

Query: 236 GLGQVNPDKSVVIGDKVIKGRNVVLATGSKVSRINIPGIESPLVLTSDDILDLREIPKSL 295
           GLGQVNPDK+V IG + IKGRNV+LATGSKVSRINIPGI+S LVLTSDDILDLRE+PKSL
Sbjct: 239 GLGQVNPDKTVTIGSQTIKGRNVILATGSKVSRINIPGIDSKLVLTSDDILDLREMPKSL 298

Query: 296 AVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILAKKGMKIKTSVG 355
           AVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKIL+KKGMKIKTSVG
Sbjct: 299 AVMGGGVVGIELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILSKKGMKIKTSVG 358
```

```
Query: 356 VSEIVEANNQLTLKLNNGEEVVADKALLSIGRVPQMNGLENLEPELEMERGRIKVNAYQE 415
           VSEIVEANNQLTLKLNNGEEVVA+KALLSIGRV QMNGLENL   LEM+R RIKVN YQE
Sbjct: 359 VSEIVEANNQLTLKLNNGEEVVAEKALLSIGRVSQMNGLENL--NLEMDRNRIKVNDYQE 416

Query: 416 TSIPGIYAPGDVNGTRMLAHAAYRMGEVAAENALGGN-KRKAHLDFTPAAVYTHPEVAMV 474
           TSIPGIYAPGDVNGT+MLAHAAYRMGEVAAENA+ GN  RKA+L +TPAAVYTHPEVAMV
Sbjct: 417 TSIPGIYAPGDVNGTKMLAHAAYRMGEVAAENAMHGNTTRKANLKYTPAAVYTHPEVAMV 476

Query: 475 GMTEEQAREQYGDILVGKNSFTGNGRAIASNEAHGFVKVIAEPKYKEILGVHIIGPAAAE 534
           G+TEEQAREQYGD+L+GKNSFTGNGRAIASNEAHGFVKVIA+ KY EILGVHIIGPAAAE
Sbjct: 477 GLTEEQAREQYGDVLIGKNSFTGNGRAIASNEAHGFVKVIADAKYHEILGVHIIGPAAAE 536

Query: 535 LINEASTIMENELTVYDVAQSIHGHPTFSEVMYEAFLDVLGEAIHNPPKRK 585
           +INEA+TIME+ELTV ++  SIHGHPTFSEVMYEAF DVLGEAIHNPPKRK
Sbjct: 537 MINEAATIMESELTVDELLLSIHGHPTFSEVMYEAFADVLGEAIHNPPKRK 587
```

SEQ ID 4264 (GBS681) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 165 (lane 2; MW 68.3kDa) and in Figure 188 (lane 10; MW 68kDa).

Purified GBS681-His is shown in Figure 240, lane 5-6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1391

A DNA sequence (GBSx1476) was identified in *S.agalactiae* <SEQ ID 4265> which encodes the amino acid sequence <SEQ ID 4266>. This protein is predicted to be dihydrolipoamide acetyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4466(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04497 GB:AP001509 dihydrolipoamide S-acetyltransferase
           [Bacillus halodurans]
 Identities = 187/462 (40%), Positives = 266/462 (57%), Gaps = 26/462 (5%)

Query: 1   MAVEIIMPKLGVDMQEGEILEWKKQVGDVVNEGDVLLEIMSDKTNMEIEAEDSGVLLKIT 60
           MA EI MPKL   MQEG +L+W K+ GD V  G+ L EIM+DK N+E+EA + G LLK
Sbjct: 1   MAKEIFMPKLSSTMQEGTLLQWFKEEGDRVEVGEPLFEIMTDKINIEVEAYEEGTLLKRY 60

Query: 61  HGNGDVVPVTETIGYIGAEGEEVTEASSSENTSVEENATQVTSEPEKVEETSEPSVPAAT 120
           +G  D +PV   IGYIG   E V     +E     E     T E      T+    P++
Sbjct: 61  YGEDDEIPVNHVIGYIGTPDESVP----TEPPGASEITASSTDEAGDHRTTAVKKAPSSD 116

Query: 121 SGEKVRATPAARKLAREMSIDLALVSGTGANGRVHREDVENFKGAQPRITPLARRIAEDQ 180
             E VRATPAAR++A+E  IDL  V G+G  GRV   DV  FK    + TPLA+++AE +
Sbjct: 117 R-ENVRATPAARRIAKEKRIDLRQVEGSGPEGRVQAVDVATFKKKGQKATPLAKKVAEVK 175

Query: 181 GVDIAEITGSGIRGKIVKNDVLAAMSPQAAEAPVETKATPTTEEKQLPEGVEVIKMSAMR 240
           GV + ++ GSG  GK+ + DV  A     A +PVE K                +K+S +R
Sbjct: 176 GVALEKVQGSGPYGKVYREDVEHAQ----AASPVEDKGNR-------------VKLSGLR 218

Query: 241 KAISKGMTNSYLTAPSFTLNYDIDMTEMMALRKKLIDPIMAKTGLKVSFTDLIGMAVVKT 300
           K ++K M +S  +AP  T+  +IDM+  + +R +L+  I  +TG ++S+T+++  AV
Sbjct: 219 KVVAKRMVDSAFSAPHVTITTEIDMSSTIKIRSQLLGMIEQETGYRLSYTEIVMKAVAHA 278
```

```
Query: 301 LMKPEHRYLNASLINDAQEIELHNFVNIGIAVGLDDGLIVPVVHNADQMSLSDFVIASKD 360
           LM   H  +NAS   +  EI  H  V+IG+AV ++ GL+VPVV + D+  L+      K
Sbjct: 279 LMS--HPTINASFFEN--EIVYHEDVHIGLAVAVEGGLVVPVVKHVDKKGLAQLTNECKT 334

Query: 361 VIKKTQEGKLKSAEMSGSTFSITNLGMFGTKTFNPIINQPNSAILGVGATIPTPTVVDGE 420
           V    ++ +L    MSG TF+I+NLGM+    F P+INQP SAILGVG     P  +DG+
Sbjct: 335 VAMAARDNRLSQEMMSGGTFTISNLGMYAIDVFTPVINQPESAILGVGRIQEKPVGIDGQ 394

Query: 421 IVARPIMAMCLTIDHRIVDGMNGAKFMVDLKNLMENPFGLLI 462
           I  RP+M   L+ DHR++DG   A F+ D+K+++E PF LL+
Sbjct: 395 IELRPMMTASLSFDHRVIDGAPAAAFLTDVKSMLEQPFQLLM 436
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4267> which encodes the amino acid sequence <SEQ ID 4268>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4774(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 354/473 (74%), Positives = 390/473 (81%), Gaps = 15/473 (3%)

Query: 1   MAVEIIMPKLGVDMQEGEILEWKKQVGDVVNEGDVLLEIMSDKTNMEIEAEDSGVLLKIT 60
           MA EIIMPKLGVDMQEGEI+EWKKQ GD VNEGD+LLEIMSDKTNME+EAEDSGVLLKIT
Sbjct: 1   MAFEIIMPKLGVDMQEGEIIEWKKQEGDTVNEGDILLEIMSDKTNMELEAEDSGVLLKIT 60

Query: 61  HGNGDVVPVTETIGYIGAEGEEVTEASSSENTS-----VEENATQVTSEPEKVEETSEPS 115
              GD VPVTE IGYIGAEGE V   +SSE T+        +A    +  E V   + P
Sbjct: 61  RQAGDTVPVTEVIGYIGAEGESVDTIASSEKTTEIPVPASADAGPAVAPKENVASPA-PQ 119

Query: 116 VPAAT----SGEKVRATPAARKLAREMSIDLALVSGTGANGRVHREDVENFKGAQPRITP 171
           V A      +G KVRATPAARK A EM IDL  V GTG  GRVH+EDVENFKGAQP+ +P
Sbjct: 120 VAATAIPQGNGGKVRATPAARKAAAEMGIDLGQVPGTGPKGRVHKEDVENFKGAQPKASP 179

Query: 172 LARRIAEDQGVDIAEITGSGIRGKIVKNDVLAAMSPQAAEAPVETKATPTTEEK--QLPE 229
           LAR+IA D+G+D+A ++G+G  GK++K D++A +    A  P E KA    EEK   LPE
Sbjct: 180 LARKIAADKGIDLATVSGTGFNGKVMKEDIMAILE---AAKPAEAKAPAAKEEKVVDLPE 236

Query: 230 GVEVIKMSAMRKAISKGMTNSYLTAPSFTLNYDIDMTEMMALRKKLIDPIMAKTGLKVSF 289
           GVE   MSAMRKAISKGMTNSYLTAP+FTLNYDIDMTEM+ALRKKLIDPIMAKTGLKVSF
Sbjct: 237 GVEHKPMSAMRKAISKGMTNSYLTAPTFTLNYDIDMTEMIALRKKLIDPIMAKTGLKVSF 296

Query: 290 TDLIGMAVVKTLMKPEHRYLNASLINDAQEIELHNFVNIGIAVGLDDGLIVPVVHNADQM 349
           TDLIGMAVVKTLMKPEH Y+NASLINDA +IELH FVN+GIAVGLDDGLIVPV+H A++M
Sbjct: 297 TDLIGMAVVKTLMKPEHEYMNASLINDANDIELHRFVNLGIAVGLDDGLIVPVIHGANKM 356

Query: 350 SLSDFVIASKDVIKKTQEGKLKSAEMSGSTFSITNLGMFGTKTFNPIINQPNSAILGVGA 409
            LSDFV+ASKDVIKK Q GKLK+AEMSGSTFSITNLGMFGTKTFNPIINQPNSAILGVGA
Sbjct: 357 CLSDFVLASKDVIKKAQTGKLKAAEMSGSTFSITNLGMFGTKTFNPIINQPNSAILGVGA 416

Query: 410 TIPTPTVVDGEIVARPIMAMCLTIDHRIVDGMNGAKFMVDLKNLMENPFGLLI 462
           TIPTPTVVDGEIV+RPIMAMCLTIDHR+VDGMNGAKFMVDLK LMENPF LLI
Sbjct: 417 TIPTPTVVDGEIVSRPIMAMCLTIDHRLVDGMNGAKFMVDLKKLMENPFELLI 469
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1392

A DNA sequence (GBSx1477) was identified in *S.agalactiae* <SEQ ID 4269> which encodes the amino acid sequence <SEQ ID 4270>. This protein is predicted to be acetoin dehydrogenase (TPP-dependent) beta chain (pdhB). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1267(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9779> which encodes amino acid sequence <SEQ ID 9780> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
           chain [Bacillus halodurans]
 Identities = 189/319 (59%), Positives = 249/319 (77%), Gaps = 1/319 (0%)

Query: 11  EAINVAMSEEMRKDEKVFLMGEDVGVYGGDFGTSVGMLEEFGAKRVRDTPISEAAIAGSA 70
           EAI  AM+ EMRK+E VF++GED+GVYGG FG + GM+EEFG++RVR+TPISEAAI+G+A
Sbjct: 8   EAIREAMTLEMRKNEDVFILGEDIGVYGGAFGVTRGMIEEFGSERVRNTPISEAAISGTA 67

Query: 71  IGAAQTGLRPIVDLTFMDFVTIAMDAIVNQGAKTNYMFGGGLSTPVTFRVASGSGIGSAA 130
           IGAA TG+RPI++L F DF+TIAMD +VNQ AK  YM+GG    P+  R  +GSG G+AA
Sbjct: 68  IGAALTGMRPILELQFSDFITIAMDNMVNQAAKLRYMYGGKAKVPMVLRTPAGSGTGAAA 127

Query: 131 QHSQSLEAWLTHIPGLKVVAPGTVNESKALLKSSILDNNPVIFLEPKALYGKKEEVNMDP 190
           QHSQSLEAW+THIPGLKVV P T  ++K LLK++I DNNPVIF E K  Y  K  V  +
Sbjct: 128 QHSQSLEAWMTHIPGLKVVQPATAYDAKGLLKAAIDDNNPVIFYEHKLCYRTKCHV-PEE 186

Query: 191 DFYIPLGKGDIKREGTDLTIVSYGRMLERVMQAAEEVAEEGINVEVVDPRTLIPLDKELI 250
           ++ IPLGK D+KR+GTD+T+V+   M+ + ++AA E+ +EGI+VEV+DPRTL+PLD+E I
Sbjct: 187 EYSIPLGKADVKRKGTDVTVVATAVMVHKALEAAVELEKEGISVEVIDPRTLVPLDEETI 246

Query: 251 IDSVKKTGKLILVNDAYKTGGFTGEIATMVAESEAFDYLDHPIVRLASEDVPVPYSRVLE 310
           I SVKKT +LI+V++A K GGF GEIA+++AESEAFDYLD PI RL  + VP+PY+  LE
Sbjct: 247 IRSVKKTSRLIVVHEAVKRGGFGGEIASIIAESEAFDYLDAPIKRLGGKPVPIPYNPTLE 306

Query: 311 QGILPDVAKIKDAIYKVVN 329
           +  +P V  I +A+ + +N
Sbjct: 307 RAAIPQVPDIIEAVKETLN 325
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4271> which encodes the amino acid sequence <SEQ ID 4272>. Analysis of this protein sequence reveals the following:

```
     Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane   81 -  97 (  81 -  97)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
           chain [Bacillus halodurans]
 Identities = 187/319 (58%), Positives = 244/319 (75%), Gaps = 1/319 (0%)
```

```
Query: 11  EAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMIEEFGPKRVKDTPISEAAISGAA 70
           EA+  AMT EMRK+E++F++GED+GVYGG FG + GMIEEFG +RV++TPISEAAISG A
Sbjct: 8   EAIREAMTLEMRKNEDVFILGEDIGVYGGAFGVTRGMIEEFGSERVRNTPISEAAISGTA 67

Query: 71  IGAAITGLRPIVDVTFMDFLTIMMDAIVNNGAKNNYMFGGGLITPVTFRVASGSGIGSAA 130
           IGAA+TG+RPI+++ F DF+TI MD +VN  AK  YM+GG    P+  R  +GSG G+AA
Sbjct: 68  IGAALTGMRPILELQFSDFITIAMDNMVNQAAKLRYMYGGKAKVPMVLRTPAGSGTGAAA 127

Query: 131 QHSQSLEAWLTHIPGIKVVAPGNANDAKGLLKSAIRDNNIVLFMEPKALYGKKEEVNQDP 190
           QHSQSLEAW+THIPG+KVV P  A DAKGLLK+AI DNN V+F E K  Y  K  V ++
Sbjct: 128 QHSQSLEAWMTHIPGLKVVQPATAYDAKGLLKAAIDDNNPVIFYEHKLCYRTKCHVPEE- 186

Query: 191 DFYIPLGKGDIKREGTDLTIVSYGRMLERVLQAAEEVAADGINVEVVDPRTLIPLDKELI 250
           ++ IPLGK D+KR+GTD+T+V+   M+ + L+AA E+  +GI+VEV+DPRTL+PLD+E I
Sbjct: 187 EYSIPLGKADVKRKGTDVTVVATAVMVHKALEAAVELEKEGISVEVIDPRTLVPLDEETI 246

Query: 251 IESVKKTGKLMLVNDAYKTGGFIGEIATMITESEAFDYLDHPIVRLASEDVPVPYARVLE 310
           I SVKKT +L++V++A K GGF GEIA++I ESEAFDYLD PI RL  + VP+PY   LE
Sbjct: 247 IRSVKKTSRLIVVHEAVKRGGFGGEIASIIAESEAFDYLDAPIKRLGGKPVPIPYNPTLE 306

Query: 311 QAILPDVEKIKAAIVKMAN 329
           +A +P V  I  A+ +  N
Sbjct: 307 RAAIPQVPDIIEAVKETLN 325
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 286/331 (86%), Positives = 310/331 (93%)

Query: 1   MSETKVMALREAINVAMSEEMRKDEKVFLMGEDVGVYGGDFGTSVGMLEEFGAKRVRDTP 60
           MSETK+MALREA+N+AM+EEMRKDE +FLMGEDVGVYGGDFGTSVGM+EEFG KRV+DTP
Sbjct: 1   MSETKLMALREAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMIEEFGPKRVKDTP 60

Query: 61  ISEAAIAGSAIGAAQTGLRPIVDLTFMDFVTIAMDAIVNQGAKTNYMFGGGLSTPVTFRV 120
           ISEAAI+G+AIGAA TGLRPIVD+TFMDF+TI MDAIVN GAK NYMFGGGL TPVTFRV
Sbjct: 61  ISEAAISGAAIGAAITGLRPIVDVTFMDFLTIMMDAIVNNGAKNNYMFGGGLITPVTFRV 120

Query: 121 ASGSGIGSAAQHSQSLEAWLTHIPGLKVVAPGTVNESKALLKSSILDNNPVIFLEPKALY 180
           ASGSGIGSAAQHSQSLEAWLTHIPG+KVVAPG  N++K LLKS+I DNN V+F+EPKALY
Sbjct: 121 ASGSGIGSAAQHSQSLEAWLTHIPGIKVVAPGNANDAKGLLKSAIRDNNIVLFMEPKALY 180

Query: 181 GKKEEVNMDPDFYIPLGKGDIKREGTDLTIVSYGRMLERVMQAAEEVAEEGINVEVVDPR 240
           GKKEEVN DPDFYIPLGKGDIKREGTDLTIVSYGRMLERV+QAAEEVA +GINVEVVDPR
Sbjct: 181 GKKEEVNQDPDFYIPLGKGDIKREGTDLTIVSYGRMLERVLQAAEEVAADGINVEVVDPR 240

Query: 241 TLIPLDKELIIDSVKKTGKLILVNDAYKTGGFTGEIATMVAESEAFDYLDHPIVRLASED 300
           TLIPLDKELII+SVKKTGKL+LVNDAYKTGGF GEIATM+ ESEAFDYLDHPIVRLASED
Sbjct: 241 TLIPLDKELIIESVKKTGKLMLVNDAYKTGGFIGEIATMITESEAFDYLDHPIVRLASED 300

Query: 301 VPVPYSRVLEQGILPDVAKIKDAIYKVVNKG 331
           VPVPY+RVLEQ ILPDV KIK AI K+ NKG
Sbjct: 301 VPVPYARVLEQAILPDVEKIKAAIVKMANKG 331
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1393

A DNA sequence (GBSx1478) was identified in *S.agalactiae* <SEQ ID 4273> which encodes the amino acid sequence <SEQ ID 4274>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.03    Transmembrane  161 - 177 ( 161 - 178)

----- Final Results -----
              bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9777> which encodes amino acid sequence <SEQ ID 9778> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04495 GB:AP001509 acetoin dehydrogenase (TPP-dependent) alpha
           chain [Bacillus halodurans]
 Identities = 148/317 (46%), Positives = 214/317 (66%), Gaps = 1/317 (0%)

Query: 8   LSKEQHLDMFLKMQRIRDVDMKFNKLVRRGFVQGMTHFSVGEEAASVGAIQDLTDSDIIF 67
           +++++ +D+F +M  IR  + K ++   +G + G TH +VG+EA++VG+I  L + D +
Sbjct: 10  MTEKKLVDLFKQMWLIRYFEEKVDEFFAKGMIHGTTHLAVGQEASAVGSIAVLEERDKLT 69

Query: 68  SNHRGHGQTIAKGIDIGGMFAELAGKATGTSKGRGGSMHLANLEKGNYGTNGIVGGGYAL 127
           S HRGHG  IAKG D+  M AEL G+ TG  KG+GGSMH+A++E+GN G NGIVGGG+++
Sbjct: 70  STHRGHGHCIAKGADVNRMMAELFGRETGYCKGKGGSMHIADVERGNLGANGIVGGGFSI 129

Query: 128 AVGAALTQQYEGTDNIVIAFSGDSATNEGSFHESVNLAAVWNLPVIFFIINNRYGISTDI 187
           A GAALT + +    +V+ F GD A+NEGSFHE+VNLA++W LPV+F   NN+YG+S  +
Sbjct: 130 ATGAALTSKMKKEGYVVLCFFGDGASNEGSFHEAVNLASIWKLPVVFICENNQYGMSGSV 189

Query: 188 TYSTKIPHLYMRADAYGIPGHYVEDGNDLMAVYEKMHEVINYVRSGNGPAIVEVESYRWF 247
                I H+  RA  YGIPG  V DGND+ AV   +   ++  R G GP IVE ++YRW
Sbjct: 190 KEMINIEHISDRAAGYGIPG-MVVDGNDVFAVMNVVGRAVDRARRGEGPTIVEAKTYRWK 248

Query: 248 GHSTADAGVYRTKEEVDSWKAKDPVKRYRAYLIENEIATEEELAAIEAQVIKEVEEGVKF 307
           GHS +DA  YRT+EE   W+ KDP+ R RA L++  I TEEE  +I+ +  +++E+ V+F
Sbjct: 249 GHSKSDAKKYRTREEEKEWREKDPIARLRATLVKEGIVTEEEADSIQEEAKQKIEDSVQF 308

Query: 308 AEESPFPDMSVAFEDVF 324
           A  SP P++    EDV+
Sbjct: 309 ARNSPEPEIESLLEDVY 325
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4275> which encodes the amino acid sequence <SEQ ID 4276>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3502(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 244/326 (74%), Positives = 278/326 (84%)

Query: 1   MEVRMVTLSKEQHLDMFLKMQRIRDVDMKFNKLVRRGFVQGMTHFSVGEEAASVGAIQDL 60
           ME  MVT+SKEQHLDMFLKM+RIR+ D + NKLVRRGFVQGMTHFSVGEEAA+VGA+  L
Sbjct: 1   MEAEMVTVSKEQHLDMFLKMERIREFDSRINKLVRRGFVQGMTHFSVGEEAANVGAVAHL 60

Query: 61  TDSDIIFSNHRGHGQTIAKGIDIGGMFAELAGKATGTSKGRGGSMHLANLEKGNYGTNGI 120
           +  DIIFSNHRGHGQ+IAK +D+  M AELAGKATG SKGRGGSMHLA+ EKGNYGTNGI
Sbjct: 61  SYDDIIFSNHRGHGQSIAKDMDLNKMMAELAGKATGVSKGRGGSMHLADFEKGNYGTNGI 120

Query: 121 VGGGYALAVGAALTQQYEGTDNIVIAFSGDSATNEGSFHESVNLAAVWNLPVIFFIINNR 180
           VGGGYALAVGAALTQQY+GT+NI +AFSGD ATNEGSFHESVN+AA W LPVIFFIINNR
Sbjct: 121 VGGGYALAVGAALTQQYKGTNNIAVAFSGDGATNEGSFHESVNMAATWKLPVIFFIINNR 180

Query: 181 YGISTDITYSTKIPHLYMRADAYGIPGHYVEDGNDLMAVYEKMHEVINYVRSGNGPAIVE 240
           YGIS  I  +T  PHLY RA+AYG+PG Y EDGND+MAVYE M + + +VR GNGPAIVE
Sbjct: 181 YGISMSINNATNTPHLYTRAEAYGVPGFYCEDGNDVMAVYETMGKAVEHVRGGNGPAIVE 240
```

```
Query: 241 VESYRWFGHSTADAGVYRTKEEVDSWKAKDPVKRYRAYLIENEIATEEELAAIEAQVIKE 300
           VESYRWFGHSTADAG YRTKEEVD WK KDP+ +YR YL    IAT++EL AI+AQV KE
Sbjct: 241 VESYRWFGHSTADAGKYRTKEEVDEWKEKDPMIKYRTYLTSEGIATDDELDAIQAQVKKE 300

Query: 301 VEEGVKFAEESPFPDMSVAFEDVFVD 326
           V++  +FA+ SP P++SVAFEDV+VD
Sbjct: 301 VDDAYEFAQNSPDPELSVAFEDVWVD 326
```

A related GBS gene <SEQ ID 8797> and protein <SEQ ID 8798> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      -14.75
GvH: Signal Score (-7.5): -4.24
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -3.03 threshold:  0.0
   INTEGRAL    Likelihood = -3.03   Transmembrane  161 - 177 ( 161 - 178)
   PERIPHERAL  Likelihood =  3.55     117
 modified ALOM score:   1.11

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01791(298 - 1278 of 1578)
EGAD|108208|BS0806(3 - 327 of 333) acetoin dehydrogenase E1 component {Bacillus subtilis}
OMNI|NT01BS0951 acetoin:DCPIP oxidoreductase alpha subunit
GP|2780395|dbj|BAA24296.1||D78509 YfjK {Bacillus subtilis}
GP|2633130|emb|CAB12635.1||Z99108 acetoin dehydrogenase E1 component (TPP-dependent alpha
subunit) {Bacillus subtilis} GP|2957146|gb|AAC05582.1||AF006075 TPP-dependent acetoin
dehydrogenase, E1 alpha-subunit {Bacillus subtilis} PIR|D69581|D69581 acetoin dehydrogenase
E1 component (TPP-dependent alpha subuni) acoA - Ba
%Match = 26.3
%Identity = 45.3  %Similarity = 65.7
Matches = 148  Mismatches = 109  Conservative Sub.s = 67

231       261       291       321       351       381       411       441
F*IEMPFTKTKKAVQILASCEKNLYNN*VIKIFLEVRMVTLSKEQHLDMFLKMQRIRDVDMKFNKLVRRGFVQGMTHFSV
                                :|:   ::|::|: | |: ||  ||  : | ::|  :| : |  |:
                              MKLLKREGLSLTEEKALWMYQKMLEIRGFEDKVHELFAQGVLPGFVHLYA
                                      10        20        30        40        50

471       501       531       561       591       621       651       681
GEEAASVGAIQDLTDSDIIFSNHRGHGQTIAKGIDIGGMFAELAGKATGTSKGRGGSMHLANLEKGNYGTNGIVGGGYAL
|||| :||    | | | | | |||||: |||| |: || ||: |||||  ||:|||||:|:|:||  | |||||||: |
GEEAVAVGVCAHLHDGDSITSTHRGHGHCIAKGCDLDGMMAEIFGKATGLCKGKGGSMHIADLDKGMLGANGIVGGGFTL
        60        70        80        90       100       110       120       130

711       741       771       801       831       861       891       921
AVGAALTQQYEGTDNIVIAFSGDSATNEGSFHESVNLAAVWNLPVIFFIINNRYGISTDITYSTKIPHLYMRADAYGIPG
| |:||| :|: | |: : | || | |:|:||| :||||||||||:|   || || :|   |::    :  || || :||
ACGSALTAKYKQTKNVSVCFFGDGANNQGTFHEGLNLAAVWNLPVVFVAENNGYGEATPFEYASACDSIADRAAAYNMPG
       140       150       160       170       180       190       200       210

951       981       1011      1041      1071      1098      1128      1158
HYVEDGNDLMAVYEKMHEVINYVRSGNGPAIVEVESYRWFGHSTADAGVYRTKEE-VDSWKAKDPVKRYRAYLIENEIAT
    || |::|||:   | |   |:| ||:::|  :|| :||   ||  |:||:| |:  : || :: :: ||::   |
-VTVDGKDILAVYQAAEEAIERARNGGGPSLIECMTYRNYGHFEGDAQTYKTKDERVEHLEEKDAIQGFKNYLLKETDAN
        220       230       240       250       260       270       280
```

```
1188      1218      1248      1278      1308      1338      1368      1398
EEELAAIEAQVIKEVEEGVKFAEESPFPDMSVAFEDVFVD*NNLK*MRFISFYYSID*KTDIRRK**AKLKLWLCAKRLM
  :|: || :| : :|: | |:|:||:|  |  : ||:|
--KLSDIEQRVSESIEKAVSFSEDSPYPKDSELLTDVYVSYEKGGM
          300       310       320       330
```

SEQ ID 8798 (GBS403) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 2; MW 64.4kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 4; MW 39.5kDa).

GBS403-GST was purified as shown in Figure 218, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1394

A DNA sequence (GBSx1479) was identified in *S.agalactiae* <SEQ ID 4277> which encodes the amino acid sequence <SEQ ID 4278>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2464(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9775> which encodes amino acid sequence <SEQ ID 9776> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12414 GB:Z99107 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 328/643 (51%), Positives = 443/643 (68%), Gaps = 9/643 (1%)

Query: 9   MIILQGNKIERSFSGDVLFDNINIQVDQRDRIALVGRNGAGKSTLLKILVGEEAPTKGEI 68
           M+ILQ N++ +SF  D + +NI ++V  RDRIA+VGRNGAGKSTLLKI+ G+ +  KGEI
Sbjct: 1   MMILQANQLSKSFGADTILNNIKLEVRNRDRIAIVGRNGAGKSTLLKIIAGQLSYEKGEI 60

Query: 69  NKKRDLSLSYLAQDSRFQSENTIFQEMLQVFDSLREVEKRLRELELQMGQVSGSDLEQLM 128
            K +D+++ YLAQ +   S+ TI +E+L VFD L+ +EK +R +E +M       +LE +M
Sbjct: 61  IKPKDITMGYLAQHTGLDSKLTIKEELLTVFDHLKAMEKEMRAMEEKMAAADPGELESIM 120

Query: 129 KTYDILSEEFREKGGFTYESDIKAILNGFKFNSDMWEMPISELSGGQNTRLALAKMLLEK 188
           KTYD L +EF++KGG+ YE+D++++L+G  F+        + LSGGQ TRLAL K+LL +
Sbjct: 121 KTYDRLQQEFKDKGGYQYEADVRSVLHGLGFSHFDDSTQVQSLSGGQKTRLALGKLLLTQ 180

Query: 189 PELLVLDEPTNHLDIDTIAWLENYLVNYQGALIIVSHDRYFLDKVATVTYDLTTHSLDRY 248
           P+LL+LDEPTNHLDIDT+ WLE+YL  Y GA++IVSHDRYFLDKV    Y+++     +Y
Sbjct: 181 PDLLILDEPTNHLDIDTLTWLEHYLQGYSGAILIVSHDRYFLDKVVNQVYEVSRAESKKY 240

Query: 249 VGNYSKFMDLKAEKIATEEKNFEKQQKEIAKLEDFVQRNIVRASTTKRAQARRKQLEKME 308
            GNYS ++D KA +   + K +EKQQ EIAKL+DFV RN+ RASTTKRAQ+RRKQLE+M+
Sbjct: 241 HGNYSAYLDQKAAQYEKDLKMYEKQQDEIAKLQDFVDRNLARASTTKRAQSRRKQLERMD 300

Query: 309 RLDKPNVEQKSANMTFHAGKVSGNVVLTLENAAIGYEG-VSLSEPIDLDVKKFDAIAIVG 367
            + KP  ++KSAN  F   K SGN VL +++  I YE    L   +   + + ++ A+VG
Sbjct: 301 VMSKPLGDEKSANFHFDITKQSGNEVLRVQDLTISYENQPPLLTEVSFMLTRGESAALVG 360

Query: 368 PNGIGKSTLIKSLVGQIPFIKGEAKLGANVETGYYDQSQSNLTKTNTVLDELWDAFSTTP 427
           PNGIGKSTL+K+L+  +   +G    G+NV  GYYDQ Q+ LT +  VLDELWD +   P
```

```
Sbjct: 361 PNGIGKSTLLKTLIDTLKPDQGTISYGSNVSVGYYDQEQAELTSSKRVLDELWDEYPGLP 420

Query: 428 EVEIRNRLGAFLFSGDDVKKSVSMLSGGERARLLLAKLSMENNNFLILDEPTNHLDIDSK 487
           E EIR  LG FLFSGDDV K V  LSGGE+ARL LAKL  ++ NFLILDEPTNHLD+DSK
Sbjct: 421 EKEIRTCLGNFLFSGDDVLKPVHSLSGGEKARLALAKLMLQKANFLILDEPTNHLDLDSK 480

Query: 488 EVLENALIEFDGTLLFVSHDRYFINRVATKVLEISDKGSTLYLGDYDYYLTKKAELEELA 547
           EVLENALI++ GTLLFVSHDRYFINR+AT+VLE+S      YLGDYDYY  KK E  EL
Sbjct: 481 EVLENALIDYPGTLLFVSHDRYFINRIATRVLELSSSHIEEYLGDYDYYTEKKTEQLELE 540

Query: 548 RLNEEEVSASKTEIDVTSD----YETQKANQKEFRKITRRVVEIEARLEVLENDENNING 603
           ++N++E    KT   V SD    YE +K  +K+ R+  RR+ EIE  ++ +E + +  +
Sbjct: 541 KMNQQE-ETDKTPATVKSDSKRSYEEEKEWKKKERQRLRRIEEIETTVQTIEENISRNDE 599

Query: 604 LMLET---NDIGKLSDLQKELESIQEEQLLLMEEWENLNMRLD 643
           L+ +     D  K+  +  + E + +E   L+ EWE L+   D
Sbjct: 600 LLCDPEVYQDHEKVQAIHADNEKLNQELESLLSEWEELSTEED 642
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4279> which encodes the amino acid sequence <SEQ ID 4280>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 473/635 (74%), Positives = 545/635 (85%), Gaps = 1/635 (0%)

Query: 9   MIILQGNKIERSFSGDVLFDNINIQVDQRDRIALVGRNGAGKSTLLKILVGEEAPTKGEI 68
           MIILQGNK+ERSFSGDVLF NI++QVD+RDRIALVG NGAGKSTLLK+LVGEE PT GE+
Sbjct: 1   MIILQGNKLERSFSGDVLFQNISLQVDERDRIALVGPNGAGKSTLLKLLVGEETPTSGEV 60

Query: 69  NKKRDLSLSYLAQDSRFQSENTIFQEMLQVFDSLREVEKRLRELELQMGQVSGSDLEQLM 128
           N K+DL+LSYLAQ+SRF+S+ TI++EML+VF++LR+ EKRLR++E+ M  VSG  L +LM
Sbjct: 61  NTKKDLTLSYLAQNSRFESDQTIYEEMLKVFEALRQDEKRLRQMEMDMATVSGQVLTRLM 120

Query: 129 KTYDILSEEFREKGGFTYESDIKAILNGFKFNSDMWEMPISELSGGQNTRLALAKMLLEK 188
             YD+L+E FR++GGFTYESDIKAILNGFKF+  MW+M I+ELSGGQNTRLALAKMLLEK
Sbjct: 121 TDYDLLTEHFRQQGGFTYESDIKAILNGFKFDESMWQMTIAELSGGQNTRLALAKMLLEK 180

Query: 189 PELLVLDEPTNHLDIDTIAWLENYLVNYQGALIIVSHDRYFLDKVATVTYDLTTHSLDRY 248
           PELLVLDEPTNHLDI+TIAWLENYL NYQGALIIVSHDRYFLDKVATVT DLT + LDRY
Sbjct: 181 PELLVLDEPTNHLDIETIAWLENYLANYQGALIIVSHDRYFLDKVATVTLDLTPNGLDRY 240

Query: 249 VGNYSKFMDLKAEKIATEEKNFEKQQKEIAKLEDFVQRNIVRASTTKRAQARRKQLEKME 308
            GNYS+FM LKAEK+  EEK F+KQQKEIAKLEDFVQ+NIVRASTTKRAQARRKQLEK+E
Sbjct: 241 SGNYSRFMALKAEKLVAEEKQFDKQQKEIAKLEDFVQKNIVRASTTKRAQARRKQLEKIE 300

Query: 309 RLDKPNVEQKSANMTFHAGKVSGNVVLTLENAAIGYEGVSLSEPIDLDVKKFDAIAIVGP 368
           RLDKP   +KSA+MTFHA K SGNVVL +E AAIGY    LSEPI++D+ K DAIA+VGP
Sbjct: 301 RLDKPTGGRKSAHMTFHAEKPSGNVVLRVEEAAIGYGDQVLSEPINVDINKLDAIAVVGP 360

Query: 369 NGIGKSTLIKSLVGQIPFIKGEAKLGANVETGYYDQSQSNLTKTNTVLDELWDAFSTTPE 428
           NGIGKSTLIKS++GQ+P +KG+ K GANVETGYYDQ+QS+LT +NTVL+ELW  FSTTPE
Sbjct: 361 NGIGKSTLIKSIIGQLPLLKGQLKYGANVETGYYDQTQSHLTSSNTVLEELWQDFSTTPE 420

Query: 429 VEIRNRLGAFLFSGDDVKKSVSMLSGGERARLLLAKLSMENNNFLILDEPTNHLDIDSKE 488
           V+IRNRLGAFLFSGDDVKKSV+MLSGGE+ARLLLAKLSMENNNFL+LDEPTNHLDIDSKE
Sbjct: 421 VDIRNRLGAFLFSGDDVKKSVAMLSGGEKARLLLAKLSMENNNFLVLDEPTNHLDIDSKE 480

Query: 489 VLENALIEFDGTLLFVSHDRYFINRVATKVLEISDKGSTLYLGDYDYYLTKKAELEELAR 548
           VLENALI+FDGTLLFVSHDRYFINR+ATKVLEI++ GSTLYLGDYDYYL KKAELEELAR
```

```
Sbjct: 481 VLENALIDFDGTLLFVSHDRYFINRLATKVLEITENGSTLYLGDYDYYLEKKAELEELAR 540

Query: 549 LNEEEVSASKTEIDVTSDYETQKANQKEFRKITRRVVEIEARLEVLENDENNINGLMLET 608
           L   E      E   T DY+ QKANQKE R++TRR  EIEARLE +E     I   M  +
Sbjct: 541 LAAGETVEETKEASAT-DYQLQKANQKERRRLTRRYEEIEARLETIEERIGAIQEDMHAS 599

Query: 609 NDIGKLSDLQKELESIQEEQLLLMEEWENLNMRLD 643
           ND  +L   QKE + + +EQ  LMEEWE +  +++
Sbjct: 600 NDTAQLIAWQKEWDQLDQEQEALMEEWETIAEQIE 634
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1395

A DNA sequence (GBSx1480) was identified in *S.agalactiae* <SEQ ID 4281> which encodes the amino acid sequence <SEQ ID 4282>. This protein is predicted to be thiophene degradation protein F (thdF). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0876(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9773> which encodes amino acid sequence <SEQ ID 9774> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4283> which encodes the amino acid sequence <SEQ ID 4284>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0795(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 384/458 (83%), Positives = 427/458 (92%)

Query: 12  MSITKEFDTIAAISTPLGEGAIGIVRISGTDALKIASKIYRGKDLSAIQSHTLNYGHIVD 71
           MSITKEFDTI AISTPLGEGAIGIVR+SGTDAL IA  +++GK+L  + SHT+NYGHI++
Sbjct: 1   MSITKEFDTITAISTPLGEGAIGIVRLSGTDALAIAQSVFKGKNLEQVASHTINYGHIIN 60

Query: 72  PDKNEILDEVMLGVMLAPKTFTREDVIEINTHGGIAVTNEILQLILRHGARMAEPGEFTK 131
           P    I+DEVM+ VMLAPKTFTRE+V+EINTHGGIAVTNEILQL++R GARMAEPGEFTK
Sbjct: 61  PKTGTIIDEVMVSVMLAPKTFTRENVVEINTHGGIAVTNEILQLLIRQGARMAEPGEFTK 120

Query: 132 RAFLNGRVDLTQAEAVMDLIRAKTDKAMDIAVKQLDGSLKTLINNTRQEILNTLAQVEVN 191
           RAFLNGRVDLTQAEAVMD+IRAKTDKAM IAVKQLDGSL  LIN+TRQEILNTLAQVEVN
Sbjct: 121 RAFLNGRVDLTQAEAVMDIIRAKTDKAMTIAVKQLDGSLSQLINDTRQEILNTLAQVEVN 180

Query: 192 IDYPEYDDVEEMTTTLMREKTQEFQALMENLLRTARRGKILREGLSTAIIGRPNVGKSSL 251
           IDYPEYDDVEEMTT L+REKTQEFQ+L+E+LLRTA+RGKILREGLSTAIIGRPNVGKSSL
Sbjct: 181 IDYPEYDDVEEMTTALLREKTQEFQSLLESLLRTAKRGKILREGLSTAIIGRPNVGKSSL 240

Query: 252 LNNLLREEKAIVTDIEGTTRDVIEEYVNIKGVPLKLVDTAGIRDTDDIVEKIGVERSKKA 311
           LNNLLRE+KAIVTDI GTTRDVIEEYVNIKGVPLKLVDTAGIR+TDD+VE+IGVERSKKA
Sbjct: 241 LNNLLREDKAIVTDIAGTTRDVIEEYVNIKGVPLKLVDTAGIRETDDLVEQIGVERSKKA 300
```

```
Query: 312 LEEADLVLLVLNSSEPLTLQDRSLLELSKESNRIVLLNKTDLPQKIEVNELPKNVIPISV 371
           L+EADLVLLVLN+SE LT QDR+LL LS++SNRI+LLNKTDL QKIE+ +LP + IPISV
Sbjct: 301 LQEADLVLLVLNASEKLTDQDRALLNLSQDSNRIILLNKTDLEQKIELEQLPDDYIPISV 360

Query: 372 LENENIDKIEERINDIFFDNAGMVEHDATYLSNARHISLIEKAVDSLKAVNEGLELGMPV 431
           L N+NI+ IE+RIN +FFDNAG+VE DATYLSNARHISLIEKAV SL+AVN+GL LGMPV
Sbjct: 361 LTNQNINLIEDRINQLFFDNAGLVEQDATYLSNARHISLIEKAVQSLEAVNDGLALGMPV 420

Query: 432 DLLQVDMTRTWEILGEITGDAAPDELITQLFSQFCLGK 469
           DLLQVD+TRTWEILGEITGDAAPDELITQLFSQFCLGK
Sbjct: 421 DLLQVDLTRTWEILGEITGDAAPDELITQLFSQFCLGK 458
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1396

A DNA sequence (GBSx1481) was identified in *S.agalactiae* <SEQ ID 4285> which encodes the amino acid sequence <SEQ ID 4286>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.18   Transmembrane  280 - 296 ( 276 - 299)
   INTEGRAL    Likelihood = -4.83   Transmembrane  249 - 265 ( 243 - 266)

----- Final Results -----
               bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD40365 GB:AF036485 hypothetical protein [Plasmid pNZ4000]
 Identities = 88/306 (28%), Positives = 149/306 (47%), Gaps = 17/306 (5%)

Query: 1   MIVEQKFGNGFTWIN---IEAEQLRTETSEIQAKY-LDSEIITYALDDYERAFMECSHIK 56
           MI  +K  NG  WI    I AE+  T    ++ +Y +D +II Y  D+ E        I
Sbjct: 1   MIKPEKTINGTKWIETIQINAEERAT----LEDQYGIDEDIIEYVTDNDESTNYVYD-IN 55

Query: 57  GKEVLTIIFNTIDLKQKESYYETVPMTFCLSHDRLITVTRSRNSYMLELLQKYLDRNPDV 116
             + L I      L +    Y T P    L    L T  +S    +   L   LD NP+V
Sbjct: 56  EDDQLFIFLAPYALDKDALRYITQPFGMLLHKGVLFTFNQSGIPEVNTALYSALD-NPEV 114

Query: 117 -SPKKFLFAALTLITKQYFNVVSKIDREKDILNRQLREQTTNKRLLAMSDLETGSVYLLT 175
            S   F+   L  +   +  +   I ++++ L++ L  +T N  L+++S L+    +L +
Sbjct: 115 KSVDAFILETLFTVVVSFIPISRAITKKRNYLDKMLNRKTKNSDLVSLSYLQQTLTFLSS 174

Query: 176 AANQNALVLEQLDVHPSQRFNSEVEKEQLS---DALIEAHQLVSMTQLNSQVLSQLSSTF 232
           A   N   L +LD  P   F    +++++    D  IE  Q+  M ++ +QV+ ++  T
Sbjct: 175 AVQTN---LSELDRLPKTHFGVGADQDKIDLFEDVQIEGEQVQRMFEIETQVVDRIDHTL 231

Query: 233 NNVLNNNLNENLTGLNIISINLAIIAAITGFFGMNIPLPLTESRSSWLIVIATSVLLWVI 292
           N++ NNNLN+ +  L I S+ +A+   I+GF+GMN+ LPL   + +W++ +  SV+L V
Sbjct: 232 NSLANNNLNDTMKFLTIWSLTMAVPTIISGFYGMNVKLPLAGMQYAWMLTLGISVVLIVA 291

Query: 293 IAQILK 298
           +  +LK
Sbjct: 292 MLIMLK 297
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1397

A DNA sequence (GBSx1482) was identified in *S.agalactiae* <SEQ ID 4287> which encodes the amino acid sequence <SEQ ID 4288>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1437(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1398

A DNA sequence (GBSx1483) was identified in *S.agalactiae* <SEQ ID 4289> which encodes the amino acid sequence <SEQ ID 4290>. This protein is predicted to be exonuclease RexA. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3165(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9771> which encodes amino acid sequence <SEQ ID 9772> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC12966 GB:U76424 exonuclease RexA [Lactococcus lactis]
 Identities = 522/1211 (43%), Positives = 747/1211 (61%), Gaps = 73/1211 (6%)

Query: 28    KRTPEQIEAIYTFGNNVLVSASAGSGKTFVMVERILDKLLRGVPIDSLFISTFTVKAAGE 87
             K TPEQ EAI++ G N+LVSASAGSGKTFVM +RI++K+ +G+ ID LFISTFT KAA E
Sbjct: 5     KLTPEQNEAIHSSGKNILVSASAGSGKTFVMAQRIVEKVKQGIEIDRLFISTFTKKAASE 64

Query: 88    LKERLEKKINESLKSAESDDLKQFLTQQLVGIQTADIGTMDAFTQKIVNQYGYTLGISPI 147
             L+ RLE+ + ++ + +  D+    LT  L  +  ADIGTMD+FTQK+       + I P
Sbjct: 65    LRMRLERDLKKARQESSDDEEAHRLTLALQNLSNADIGTMDSFTQKLTKANFNRVNIDPN 124

Query: 148   FRILQDKNEQDVIKNEVYADLFSDYMTGKNAAS-----FIKLVKNFSGNRKDSKAFREMV 202
             FRIL D+ E D+I+ EV+  L   Y++   + +     F KL+KNFS +R +   F+++V
Sbjct: 125   FRILADQTESDLIRQEVFEQLVESYLSADESLNISKDKFEKLIKNFSKDR-NILGFQKVV 183

Query: 203   YKVYAFSQSTDNPKRWMQTVFLKGAQTYTDFEAIPDQEVSSLLNVMQT--TANQLRDLTD 260
             Y +Y F+ +T+NP  W++  FLKG +TY   +++ D      +NV +   T  +L +
Sbjct: 184   YTIYRFASATENPISWLENQFLKGFETY---KSLTDLSEDFTVNVKENLLTFFELLEAIS 240

Query: 261   QEDYKQLTAKGVPTANYKKHLKIIENL-VHWSQDFNLLYGKKGLTNLARDITNVIPSGND 319
             ++D+   TA           L I  ++ V  S+D  L   KK  +   +D+
Sbjct: 241   KKDFVTCTAL---------FLSIDTDIRVGSSKDEALSALKKDFSAQKQDL--------- 282

Query: 320   VTVAGVKYPIFKQLHNRIVGLKHLEVIFKYQGESLFLLELLQSFVLDFSEQYLQEKIQEN 379
```

```
              V     P   +L   +  +KH ++I KYQ ++  +   LQ F++DF + YL+ K  EN
Sbjct: 283  --VGSKSKP--GELRKFVDKIKHGQLIEKYQNQAFEIASDLQKFIIDFYKTYLERKKNEN 338

Query: 380  AFEFSDIAHFAIQILEENHDIRQLYQDKYHEVMVDEYQDNNHTQERMLELLSNGHNRFMV 439
            AFE+SDIAHFAI+ILEEN DIR+  ++ Y E+M+DEYQD +HTQERMLELLSNGHN FMV
Sbjct: 339  AFEYSDIAHFAIEILEENPDIRENLREHYDEIMIDEYQDTSHTQERMLELLSNGHNLFMV 398

Query: 440  GDIKQSIYRFRQADPQIFNDKYKAYQDNPSQGKLIILKENFRSQSEVLDSTNSVFTHLMD 499
            GDIKQSIY FR ADP +F +KYK+Y    +  +LI LKENFRS+ EVL+ TN +F HLMD
Sbjct: 399  GDIKQSIYGFRLADPGLFLEKYKSYDQAENPNQLIRLKENFRSRGEVLNFTNDIFKHLMD 458

Query: 500  EEVGDILYDESHQLKAGS----PRQQERHPNNKTQVLLLDTDEDDIDDSDSQQYDISPAE 555
            E++G++ Y +   L  G+    P + E+    +  +   +T E++I+DS+ +   IS  E
Sbjct: 459  EKLGEMTYGKEEALVQGNISDYPVEAEKDFYPELLLYKENTSEEEIEDSEVK---ISDGE 515

Query: 556  AKLVAKEIIRLHKEENVPFQDITLLVSSRTRNDGILQTFDRYGIPLVTDGGEQNYLKSVE 615
             K  A+EI +L  E  V  +DI +LV S++ N+ I      Y IP+V D G  ++LKS+E
Sbjct: 516  IKGAAQEIKKL-IEYGVEPKDIAILVRSKSNNNKIEDILLSYDIPVVLDEGRVDFLKSME 574

Query: 616  VMVMLDTLRSIDNPLNDYALVALLRSPMFGFNEDDLTRIAIQDVK-MAFYHKVKLSYHKE 674
            V++MLD LR+IDNPL D +LVA+LRSP+FGFNED+LTRI++Q  + + F+ K+ LS  KE
Sbjct: 575  VLIMLDVLRAIDNPLYDLSLVAMLRSPLFGFNEDELTRISVQGSRDLRFWDKILLSLKKE 634

Query: 675  GHHSDLITPELSSKIDHFMKTFQTWRDFAKWHSLYDLIWKIYNDRFYYDYVGALPKAEQR 734
            G + +LI   L  K+  F + F  WR       ++ L+WKIY + +Y+DYVGAL   E R
Sbjct: 635  GKNPELINLSLEQKLKAFNQKFTEWRKLVNKIPIHRLLWKIYTETYYFDYVGALKNGEMR 694

Query: 735  QANLYALALRANQFEKTGFKGLSRFIRMIDKVLENENDLADVEVALPQNAVNLMTIHKSK 794
            QANL AL++RA  +E +G+KGL +F+R+I+K +E  NDLA V + LPQNAV +MT HKSK
Sbjct: 695  QANLQALSVRAESYESSGYKGLFKFVRLINKFMEQNNDLASVNIKLPQNAVRVMTFHKSK 754

Query: 795  GLEFKYVFILNIDKKFSMVDITSPLILSRNQGIGIKYVADMRHELEE-EILPAVKVSMET 853
            GLEF YVF++N+  +F+  D+   +ILSR  G+G+KY+AD++ E +     P   V MET
Sbjct: 755  GLEFDYVFLMNLQSRFNDRDLKEDVILSREHGLGMKYIADLKAEPDVITDFPYALVKMET 814

Query: 854  LPYQLNKRELRLATLSEQMRLLYVAMTRAEKKLYLVGKASQT---KWADHYDLVS-ENNH 909
             PY +NK   + A LSE+MR+LYVA TRA+KKLYLVGK   T      + YD  + E
Sbjct: 815  FPYMVNKDLKQRAALSEEMRVLYVAFTRAKKKLYLVGKIKDTDKKAGLELYDAATLEGKI 874

Query: 910  LPLASRETFVTFQDWLLAVHETYKKQELFYDINFVSLEELTDHHIGMVNPSLPFNPDNK- 968
            L    R +   FQ W+LA+    K   L   +N  + +EL    +   +     PD K
Sbjct: 875  LSDKFRNSSRGFQHWILALQNATK---LPMKLNVYTKDELETEKLEFTS-----QPDFKK 926

Query: 969  -VENRQSEDIVRAIS--VLESVEQINQTY--KAAIELPTVRTPSQVKK-IYEPILDIEGV 1022
             VE  +  D + + S  + E+ + +N  Y  +AA EL +++TPSQVKK  YE  L +  V
Sbjct: 927  LVEESEKFDNIMSFSDEIKEAQKIMNYQYPHQAATELSSIQTPSQVKKRSYEKQLQVGEV 986

Query: 1023 D-VMETITKTSVDFKLPDFSTSKKQDPAALGSAVHELMQRIEMSSHVKMEDIQKALTEVN 1081
              V E +   ++DF   DF   KK   A +GSA H  MQ  +  S   +   Q  L E+
Sbjct: 987  QPVSEFVRVKNLDFS--DFG-PKKITAAEMGSATHSFMQYADF-SQADLFSFQATLDEMG 1042

Query: 1082 AETSVKAAIQIEKINYFFQETSLGKYIQEEVEHLHREAPFAMLKEDPESGEKFVVRGIID 1141
             +  +K  I I KI   F +T  G+++ E V+   +EAPF+ML+ D  + E+++VRGI D
Sbjct: 1043 FDEKIKNQIDITKILTLF-DTEFGQFLSENVDKTVKEAPFSMLRTDEFAKEQYIVRGICD 1101

Query: 1142 GYLLLENRIILFDYKTDKFVNP---LELKERYQGQMALYAEALKKSYEIEKIDKYLILLG 1198
            G++ L ++IILFDYKTD+F N     E+KERY+ QM LY+EAL+K+Y + +IDKYLILLG
Sbjct: 1102 GFVKLADKIILFDYKTDRFTNVSAISEIKERYKDQMNLYSEALQKAYHVNQIDKYLILLG 1161

Query: 1199 G-KQLEVVKMD 1208
            G +++ V K+D
Sbjct: 1162 GPRKVFVEKID 1172
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4291> which encodes the amino acid sequence <SEQ ID 4292>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
```

```
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC12966 GB:U76424 exonuclease RexA [Lactococcus lactis]
 Identities = 478/1206 (39%), Positives = 700/1206 (57%), Gaps = 65/1206 (5%)

Query: 40   KRTAQQIEAIYTSGQNILVSASAGSGKTFVMVERILDKILRGVSIDRLFISTFTVKAATE 99
            K T +Q EAI++SG+NILVSASAGSGKTFVM +RI++K+ +G+ IDRLFISTFT KAA+E
Sbjct: 5    KLTPEQNEAIHSSGKNILVSASAGSGKTFVMAQRIVEKVKQGIEIDRLFISTFTKKAASE 64

Query: 100  LRERIENKLYSQIAQTTDFQMKVYLTEQLQSLCQADIGTMDAFAQKVVSRYGYSIGISSQ 159
            LR R+E  L     +++D +    LT  LQ+L  ADIGTMD+F QK+       + I
Sbjct: 65   LRMRLERDLKKARQESSDDEEAHRLTLALQNLSNADIGTMDSFTQKLTKANFNRVNIDPN 124

Query: 160  FRIMQDKAEQDVLKQEVFSKLFNEFMNQKEA-----PVFRALVKNFSGNCKDTSAFRELV 214
            FRI+ D+ E D+++QEVF +L   +++  E+       F  L+KNFS + ++   F+++V
Sbjct: 125  FRILADQTESDLIRQEVFEQLVESYLSADESLNISKDKFEKLIKNFSKD-RNILGFQKVV 183

Query: 215  YTCYSFSQSTENPKIWLQENFLSAAKTYQRLEDIPDHDIELLLLAMQDTANQLRDVTDME 274
            YT Y F+ +TENP  WL+  FL   +TY+ L D+ + D  + +     T  +L +    +
Sbjct: 184  YTIYRFASATENPISWLENQFLKGFETYKSLTDLSE-DFTVNVKENLLTFFELLEAISKK 242

Query: 275  DYGQLTKAG-SRSAKYTKHLTIIEKLSDWVRDFKCLYGKAGLDRLIRDVTGLIPSGNDVT 333
            D+   T    S         +  E LS   +DF                        D+
Sbjct: 243  DFVTCTALFLSIDTDIRVGSSKDEALSALKKDFSA-------------------QKQDLV 283

Query: 334  VSKVKYPVFKTLHQKLKQFRHLETILMYQKDCFSLLEQLQDFVLAFSEAYLAVKIQESAF 393
             SK K    +    K+K   H + I  YQ   F +   LQ F++ F + YL  K  E+AF
Sbjct: 284  GSKSKPGELRKFVDKIK---HGQLIEKYQNQAFEIASDLQKFIIDFYKTYLERKKNENAF 340

Query: 394  EFSDIAHFAIKILEENTDIRQSYQQHYHEVMVDEYQDNNHMQERLLTLLSNGHNRFMVGD 453
            E+SDIAHFAI+ILEEN DIR++ ++HY E+M+DEYQD +H QER+L LLSNGHN FMVGD
Sbjct: 341  EYSDIAHFAIEILEENPDIRENLREHYDEIMIDEYQDTSHTQERMLELLSNGHNLFMVGD 400

Query: 454  IKQSIYRFRQADPQIFNQKFRDYQKKPEQGKVILLKENFRSQSEVLNVSNAVFSHLMDES 513
            IKQSIY FR ADP +F +K++ Y +     ++I LKENFRS+ EVLN +N +F HLMDE
Sbjct: 401  IKQSIYGFRLADPGLFLEKYKSYDQAENPNQLIRLKENFRSRGEVLNFTNDIFKHLMDEK 460

Query: 514  VGDVLYDEQHQLIAG--SHAQTVPYLDRRAQLLLYNSDKDDGNAPSDSEGISFSEVTIVA 571
            +G++ Y ++  L+ G  S        D   +LLLY  +  +         IS  E+   A
Sbjct: 461  LGEMTYGKEEALVQGNISDYPVEAEKDFYPELLLYKENTSEEEIEDSEVKISDGEIKGAA 520

Query: 572  KEIIKLHNDKGVPFEDITLLVSSRTRNDIISHTFNQYGIPIATDGGQQNYLKSVEVMVML 631
            +EI KL  + GV  +DI +LV S++ N+ I      Y IP+  D G+ ++LKS+EV++ML
Sbjct: 521  QEIKKL-IEYGVEPKDIAILVRSKSNNNKIEDILLSYDIPVVLDEGRVDFLKSMEVLIML 579

Query: 632  DTLRTINNPRNDYALVALLRSPMFAFDEDDLARIALQKDNELDKDCLYDKIQRAVIGRGA 691
            D LR I+NP  D +LVA+LRSP+F F+ED+L RI++Q   +L     +DKI  ++   G
Sbjct: 580  DVLRAIDNPLYDLSLVAMLRSPLFGFNEDELTRISVQGSRDLR---FWDKILLSLKKEGK 636

Query: 692  HPELIHDTLLGKLNVFLKTLKSWRRYAKLGSLYDLIWKIFNDRFYFDFVASQAKAEQAQA 751
            +PELI+ +L  KL  F +    WR+      ++ L+WKI+ + +YFD+V +    E  QA
Sbjct: 637  NPELINLSLEQKLKAFNQKFTEWRKLVNKIPIHRLLWKIYTETYYFDYVGALKNGEMRQA 696

Query: 752  NLYALALRANQFEKSGYKGLYRFIKMIDKVLETQNDLADVEVATPKQAVNLMTIHKSKGL 811
            NL AL++RA  +E SGYKGL++F+++I+K +E  NDLA V +  P+ AV +MT HKSKGL
Sbjct: 697  NLQALSVRAESYESSGYKGLFKFVRLINKFMEQNNDLASVNIKLPQNAVRVMTFHKSKGL 756

Query: 812  QFPYVFILNCDKRFSMTDIHKSFILNRQHGIGIKYLADIKGLLGE-TTLNSVKVSMETLP 870
            +F YVF++N   RF+  D+ +  IL+R+HG+G+KY+AD+K      T      V MET P
Sbjct: 757  EFDYVFLMNLQSRFNDRDLKEDVILSREHGLGMKYIADLKAEPDVITDFPYALVKMETFP 816
```

```
Query: 871  YQLNKQELRLATLSEEMRLLYVAMTRAEKKVYFIGK---ASKSKSQEITDPKKL-GKLLP 926
            Y +NK   + A LSEEMR+LYVA TRA+KK+Y +GK     K    E+ D   L GK+L
Sbjct: 817  YMVNKDLKQRAALSEEMRVLYVAFTRAKKKLYLVGKIKDTDKKAGLELYDAATLEGKILS 876

Query: 927  LALREQLLTFQDWLLAIADIFSTEDLYFDVRFIEDSDLTQESVGRLQTP---QLLNPDDL 983
               R     FQ W+LA+    +   L   +      +L  E +     P   +L+   +
Sbjct: 877  DKFRNSSRGFQHWILALQ---NATKLPMKLNVYTKDELETEKLEFTSQPDFKKLVEESEK 933

Query: 984  KDNRQSETIARALDMLEAVSQLNANY--EAAIHLPTVRTPSQL-KATYEPLLEPIGVDII 1040
             DN  S +     ++ EA   +N  Y  +AA  L +++TPSQ+ K +YE  L+   V  +
Sbjct: 934  FDNIMSFSD----EIKEAQKIMNYQYPHQAATELSSIQTPSQVKKRSYEKQLQVGEVQPV 989

Query: 1041 EKSSRSLSDFTLPHFSKKAKVEASHIGSALHQLMQVLPLSKP--INQQTLLDALRGIDSN 1098
             +  R + +     F  K K+ A+ +GSA H  MQ    S+    + Q  LD + G D
Sbjct: 990  SEFVR-VKNLDFSDFGPK-KITAAEMGSATHSFMQYADFSQADLFSFQATLDEM-GFD-- 1044

Query: 1099 EEVKTALDLKKIESFFCDTSLGQFFQTYQKHLYREAPFAILKLDPISQEEYVLRGIIDAY 1158
            E++K  +D+ KI + F DT  GQF         +EAPF++L+ D  ++E+Y++RGI D +
Sbjct: 1045 EKIKNQIDITKILTLF-DTEFGQFLSENVDKTVKEAPFSMLRTDEFAKEQYIVRGICDGF 1103

Query: 1159 FLFDDHIVLVDYKTDKYKQP---IELKKRYQQQLELYAEALTQTYKLPVTKRYLVLMGGG 1215
                D I+L DYKTD++       E+K+RY+ Q+ LY+EAL + Y +    +YL+L+GG
Sbjct: 1104 VKLADKIILFDYKTDRFTNVSAISEIKERYKDQMNLYSEALQKAYHVNQIDKYLILLGGP 1163

Query: 1216 KPEIVE 1221
            +   VE
Sbjct: 1164 RKVFVE 1169
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 728/1211 (60%), Positives = 916/1211 (75%), Gaps = 5/1211 (0%)

Query: 1    MMTFKPFLNPEDIAVIQTEEKNSDKKQKRTPEQIEAIYTFGNNVLVSASAGSGKTFVMVE 60
            +++F PFL+PE I  +Q  E+  D+ QKRT +QIEAIYT G N+LVSASAGSGKTFVMVE
Sbjct: 13   VISFAPFLSPEAIKHLQENERCRDQSQKRTAQQIEAIYTSGQNILVSASAGSGKTFVMVE 72

Query: 61   RILDKLLRGVPIDSLFISTFTVKAAGELKERLEKKINESLKSAESDDLKQFLTQQLVGIQ 120
            RILDK+LRGV ID LFISTFTVKAA EL+ER+E K+   +       +K +LT+QL  +
Sbjct: 73   RILDKILRGVSIDRLFISTFTVKAATELRERIENKLYSQIAQTTDFQMKVYLTEQLQSLC 132

Query: 121  TADIGTMDAFTQKIVNQYGYTLGISPIFRILQDKNEQDVIKNEVYADLFSDYMTGKNAAS 180
             ADIGTMDAF QK+V++YGY++GIS  FRI+QDK EQDV+K EV++ LF+++M  K A
Sbjct: 133  QADIGTMDAFAQKVVSRYGYSIGISSQFRIMQDKAEQDVLKQEVFSKLFNEFMNQKEAPV 192

Query: 181  FIKLVKNFSGNRKDSKAFREMVYKVYAFSQSTDNPKRWMQTVFLKGAQTYTDFEAIPDQE 240
            F  LVKNFSGN KD+ AFRE+VY  Y+FSQST+NPK W+Q  FL  A+TY   E IPD +
Sbjct: 193  FRALVKNFSGNCKDTSAFRELVYTCYSFSQSTENPKIWLQENFLSAAKTYQRLEDIPDHD 252

Query: 241  VSSLLNVMQTTANQLRDLTDQEDYKQLTAKGVPTANYKKHLKIIENLVHWSQDFNLLYGK 300
            +  LL  MQ TANQLRD+TD EDY QLT  G  +A Y KHL IIE L  W +DF  LYGK
Sbjct: 253  IELLLLAMQDTANQLRDVTDMEDYGQLTKAGSRSAKYTKHLTIIEKLSDWVRDFKCLYGK 312

Query: 301  KGLTNLARDITNVIPSGNDVTVAGVKYPIFKQLHNRIVGLKHLEVIFKYQGESLFLLELL 360
             GL  L RD+T +IPSGNDVTV+ VKYP+FK LH ++   +HLE I  YQ +   LLE L
Sbjct: 313  AGLDRLIRDVTGLIPSGNDVTVSKVKYPVFKTLHQKLKQFRHLETILMYQKDCFSLLEQL 372

Query: 361  QSFVLDFSEQYLQEKIQENAFEFSDIAHFAIQILEENHDIRQLYQDKYHEVMVDEYQDNN 420
            Q FVL FSE YL  KIQE+AFEFSDIAHFAI+ILEEN DIRQ YQ  YHEVMVDEYQDNN
Sbjct: 373  QDFVLAFSEAYLAVKIQESAFEFSDIAHFAIKILEENTDIRQSYQQHYHEVMVDEYQDNN 432

Query: 421  HTQERMLELLSNGHNRFMVGDIKQSIYRFRQADPQIFNDKYKAYQDNPSQGKLIILKENF 480
            H QER+L LLSNGHNRFMVGDIKQSIYRFRQADPQIFN K++ YQ  P QGK+I+LKENF
Sbjct: 433  HMQERLLTLLSNGHNRFMVGDIKQSIYRFRQADPQIFNQKFRDYQKKPEQGKVILLKENF 492

Query: 481  RSQSEVLDSTNSVFTHLMDEEVGDILYDESHQLKAGSPRQQERHPNNKTQVLLLDTDEDD 540
            RSQSEVL+ +N+VF+HLMDE VGD+LYDE HQL AGS  Q   + + + Q+LL ++D+DD
Sbjct: 493  RSQSEVLNVSNAVFSHLMDESVGDVLYDEQHQLIAGSHAQTVPYLDRRAQLLLYNSDKDD 552
```

```
Query: 541  IDDSDSQQYDISPAEAKLVAKEIIRLHKEENVPFQDITLLVSSRTRNDGILQTFDRYGIP 600
              ++ S    IS +E  +VAKEII+LH ++ VPF+DITLLVSSRTRND I  TF++YGIP
Sbjct: 553  -GNAPSDSEGISFSEVTIVAKEIIKLHNDKGVPFEDITLLVSSRTRNDIISHTFNQYGIP 611

Query: 601  LVTDGGEQNYLKSVEVMVMLDTLRSIDNPLNDYALVALLRSPMFGFNEDDLTRIAIQD-- 658
            + TDGG+QNYLKSVEVMVMLDTLR+I+NP NDYALVALLRSPMF F+EDDL RIA+Q
Sbjct: 612  IATDGGQQNYLKSVEVMVMLDTLRTINNPRNDYALVALLRSPMFAFDEDDLARIALQKDN 671

Query: 659  --VKMAFYHKVKLSYHKEGHHSDLITPELSSKIDHFMKTFQTWRDFAKWHSLYDLIWKIY 716
               K   Y K++ +    G H +LI   L  K++ F+KT ++WR +AK  SLYDLIWKI+
Sbjct: 672  ELDKDCLYDKIQRAVIGRGAHPELIHDTLLGKLNVFLKTLKSWRRYAKLGSLYDLIWKIF 731

Query: 717  NDRFYYDYVGALPKAEQRQANLYALALRANQFEKTGFKGLSRFIRMIDKVLENENDLADV 776
            NDRFY+D+V +  KAEQ QANLYALALRANQFEK+G+KGL RFI+MIDKVLE +NDLADV
Sbjct: 732  NDRFYFDFVASQAKAEQAQANLYALALRANQFEKSGYKGLYRFIKMIDKVLETQNDLADV 791

Query: 777  EVALPQNAVNLMTIHKSKGLEFKYVFILNIDKKFSMVDITSPLILSRNQGIGIKYVADMR 836
            EVA P+ AVNLMTIHKSKGL+F YVFILN DK+FSM DI    IL+R  GIGIKY+AD++
Sbjct: 792  EVATPKQAVNLMTIHKSKGLQFPYVFILNCDKRFSMTDIHKSFILNRQHGIGIKYLADIK 851

Query: 837  HELEEEILPAVKVSMETLPYQLNKRELRLATLSEQMRLLYVAMTRAEKKLYLVGKASQTK 896
              L E  L +VKVSMETLPYQLNK+ELRLATLSE+MRLLYVAMTRAEKK+Y +GKAS++K
Sbjct: 852  GLLGETTLNSVKVSMETLPYQLNKQELRLATLSEEMRLLYVAMTRAEKKVYFIGKASKSK 911

Query: 897  WADHYDLVSENNHLPLASRETFVTFQDWLLAVHETYKKQELFYDINFVSLEELTDHHIGM 956
              +  D       LPLA RE  +TFQDWLLA+ + +  ++L++D+ F+   +LT   +G
Sbjct: 912  SQEITDPKKLGKLLPLALREQLLTFQDWLLAIADIFSTEDLYFDVRFIEDSDLTQESVGR 971

Query: 957  VNPSLPFNPDNKVENRQSEDIVRAISVLESVEQINQTYKAAIELPTVRTPSQVKKIYEPI 1016
            +      NPD+  +NRQSE I RA+ +LE+V Q+N  Y+AAI LPTVRTPSQ+K  YEP+
Sbjct: 972  LQTPQLLNPDDLKDNRQSETIARALDMLEAVSQLNANYEAAIHLPTVRTPSQLKATYEPL 1031

Query: 1017 LDIEGVDVMETITKTSVDFKLPDFSTSKKQDPAALGSAVHELMQRIEMSSHVKMEDIQKA 1076
            L+  GVD++E  +++  DF LP FS   K + + +GSA+H+LMQ + +S  +  + +  A
Sbjct: 1032 LEPIGVDIIEKSSRSLSDFTLPHFSKKAKVEASHIGSALHQLMQVLPLSKPINQQTLLDA 1091

Query: 1077 LTEVNAETSVKAAIQIEKINYFFQETSLGKYIQEEVEHLHREAPFAMLKEDPESGEKFVV 1136
            L  +++   VK A+ ++KI  FF +TSLG++ Q   +HL+REAPFA+LK DP S E++V+
Sbjct: 1092 LRGIDSNEEVKTALDLKKIESFFCDTSLGQFFQTYQKHLYREAPFAILKLDPISQEEYVL 1151

Query: 1137 RGIIDGYLLLENRIILFDYKTDKFVNPLELKERYQGQMALYAEALKKSYEIEKIDKYLIL 1196
            RGIID Y L ++ I+L DYKTDK+  P+ELK+RYQ Q+ LYAEAL ++Y++    +YL+L
Sbjct: 1152 RGIIDAYFLFDDHIVLVDYKTDKYKQPIELKKRYQQQLELYAEALTQTYKLPVTKRYLVL 1211

Query: 1197 LGGKQLEVVKM 1207
            +GG + E+V++
Sbjct: 1212 MGGGKPEIVEV 1222
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1399

A DNA sequence (GBSx1484) was identified in *S.agalactiae* <SEQ ID 4293> which encodes the amino acid sequence <SEQ ID 4294>. This protein is predicted to be exonuclease RexB. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0660(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC12965 GB:U76424 exonuclease RexB [Lactococcus lactis]
 Identities = 363/1093 (33%), Positives = 604/1093 (55%), Gaps = 67/1093 (6%)

Query: 1    MKLLYTDINHDMTEILVNQAAHAAEAGWRIFYIAPNSLSFEKERAVLENLPQ---EASFA 57
            M++LYT+I  D+TE L+  A    E   +++YI P+S+SFEKE+ +LE L +    A F
Sbjct: 1    MEILYTEITQDLTEGLLEIALEELEKNRKVYYIVPSSMSFEKEKEILERLAKGSDTAVFD 60

Query: 58   ITITRFAQLARYFTLNQP-NQKESLNDIGLAMIFYRALASFEDGQLKVFGRLKQDASFIS 116
            + +TRF QL  YF   +    K  L  +GL+M+F R L SF+  ++ ++  L+  A F+
Sbjct: 61   LLVTRFKQLPYYFDKREKATMKTELGTVGLSMLFRRVLRSFKKDEIPLYFSLQDSAGFLE 120

Query: 117  QLVDLYKELQTANLSILELKYLHSPEKFEDLLAIFLVVSDLLREGEYDNQSKIAFFTEQV 176
             L+ L  EL TANLS+  L      ++ + +LA F     +    EY N S+   FT ++
Sbjct: 121  MLIQLRAELLTANLSVENLPDNPKNQELKKILAKFEAELSV----EYANYSEFGDFTNRL 176

Query: 177  RSGQLDVDLKNTILIVDGFTRFSAEEEALIKSLSSRCQEIIIGAYASQKAYKANFTNGNI 236
              G+ D  LK+  +I+DG+TRFSAEEE  I+S+  +    ++G Y+ + +  A   +  I
Sbjct: 177  VDGEFDQQLKDVTIIIDGYTRFSAEEELFIESIQEKVARFVVGTYSDENSLTAG--SETI 234

Query: 237  YSAGVDFLRYLATTFQTKPEFILSKWESKSGFEMISK-----NIEGKHDFTNSSHILDDT 291
            Y      +    T F+ K    L K  S +  E+ SK     +++ +   T+    L
Sbjct: 235  YVGTSQMI----TRFRNKFPVELRKIASSAVNEVYSKLTRILDLDSRFVITDEKIELKAE 290

Query: 292  AKDCITIWECINQKDEVEHVARAIRQKLYQGYRYKDILVLLGDVDSYKLQLSKIFEQYDI 351
             +    IWE  NQK E+E VA+ IRQK+ QG  +KD  VL+GD  +Y++ L ++F+ Y+I
Sbjct: 291  DEKYFRIWEAENQKVEIERVAKEIRQKIIQGAFFKDFTVLVGDPAAYEITLKEVFDLYEI 350

Query: 352  PYYFGKAETMAAHPLVHFMDSLSRIKRYRFRAEDVLNLFKTGIYGEISQDD--LDYFEAY 409
            P+++ + E+M+ HPLV F +SL  IK+  +R +DV+NL K+ +Y + + D+  +DYFE Y
Sbjct: 351  PFFYAQEESMSQHPLVIFFESLFAIKKNNYRTDDVVNLLKSKVYTDANLDEEVIDYFEYY 410

Query: 410  ISYADIKGPKKFFTDFVVGAKKFDLGRLNTIRQSLL---TPLESFV-KTKKQDGIKTLNQ 465
            +    I G KKF  +F+  ++   +  +N +R+ LL   +PL+ F+   +K+ G K ++
Sbjct: 411  VQKYKISGRKKFTEEFIE-SEFSQIELVNEMREKLLGSESPLQVFLGNNRKKTGKKWVSD 469

Query: 466  FMFFLTQVGLSDNLSRLVGQMS-ENEQE---KHQEVWKTFTDILEQFQTIFGQEKLNLDE 521
                L    +  N++        +NE +   KH++VW+     L +F  +F  EKL   E
Sbjct: 470  LQGLLENGNVMTNMNAYFSAAELQNEHQMADKHEQVWQMLISTLNEFLAVFSDEKLKSVE 529

Query: 522  FLSLLNSGMMQAEYRMVPATVDVVTVKSYDLVEPHSNQFVYALGMTQSHFPKIAQNKSLI 581
            FL +L +G+  A+YR +PA VDVV VK Y+LVEP +N+++YA+G++Q++FP+I +N +L+
Sbjct: 530  FLDILLAGLKNAKYRQIPANVDVVNVKDYELVEPKTNKYIYAIGLSQTNFPRIKKNSTLL 589

Query: 582  SDIERQLINDANDTDGHFDIMTQENLKKNHFAALSLFNAAKQELVLTIPQLLNESEDQMS 641
            SD ER  IN   D +   + +   N +KN F  LSL N+AK+ LVL++PQ++   + + S
Sbjct: 590  SDEERLEINQTTDENQFIEQLNVANYQKNQFTVLSLINSAKESLVLSMPQIMANEQGEFS 649

Query: 642  P-YLVELRDIGVPFNHKGR-QSLKEEADNIGNYKALLSRVVDLYRSAIDKEMTKEE-QTF 698
            P + + L+D       K +  +L E  ++IGN +++++ +  + R  ++ E T E+ + F
Sbjct: 650  PVFQLFLKDADEKILQKIQGVNLFESLEHIGNSRSVIAMIGQIERELVESEETSEDKRVF 709

Query: 699  WSVAVRYLRRQLTSKGIEIPIITDSLDTVTVSSDVMTRRFPEDDPLKLSSSALTTFYNNQ 758
            WS   R L +        +  +   +DTV ++ D + + +   D +  S S+   FYN +
Sbjct: 710  WSSIFRILVKSNADFQKILLDLAKDIDTVNLAPDTLEQIY--GDKIYASVSSFERFYNCE 767

Query: 759  YKYFLQYVLGLEEQDSIHPDMRHHGTYLHRVFEILMKNQGI--ESFEEKLNSAINKTNQE 816
            Y+YFL+  L LE  ++I  + +  G + H VFE +MK   +  E+F+EKL   + + ++
Sbjct: 768  YQYFLENTLSLETFENIDINSKIVGNFFHEVFEKVMKETDLSAENFDEKLTLVLQEVDKN 827

Query: 817  DVFKSLYSEDAESRYSLEILEDIARATATILR----QDSQMTVESE-------EERFELM 865
              +   +++DA +R++   LE+I R TAT+L+     D   T+ +E        E
Sbjct: 828  --YSRYFTQDATARFTWSNLEEIVRQTATVLKATVSTDELKTLLTESSFGLPKSELGNFS 885

Query: 866  IDNTIKINGIIDRIDRLSDGSLGVVDYKSSAQKFDIQKFYNGLSPQLVTYIDAISRDKEV 925
            +D+ I + G IDR+D+LS   LG +DYKSSA  F +Q+ Y+GLS Q +TY+D I   K+
Sbjct: 886  VDD-IYLRGRIDRLDQLSTDYLGAIDYKSSAHSFKLQEAYDGLSLQFMTYLDVI---KQA 941
```

```
Query: 926  EQKPPIFGAMYLHMQEPRQDLSKIKNLDDLVTKNHQALTYKGLFSEAEKEFLANGKYHL- 984
                 I+GA+YL  +    +LS+I  L ++      +++ Y+GL  E   E +  G  ++
Sbjct: 942  FPNQKIWGALYLQFKNQPINLSEINQLSEIANILKESMRYEGLVLEDAAEQI-KGIENIA 1000

Query: 985  --KDSLYSETEIAILQAHNQSLYKKASETIKSGKFLINPYTEDAKTVDGD---------Q 1033
              K ++Y+E E   L   N+  Y+ A + +K GK  INP  + ++ +D
Sbjct: 1001 LKKTNIYNEEEFEQLLKLNEEHYRAAGQRLKKGKIAINPIMKRSEGIDQSGNVRGCRYCP 1060

Query: 1034 FKSITGFEADRHM 1046
             KSI  FEA+ HM
Sbjct: 1061 LKSICRFEANIHM 1073
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4295> which encodes the amino acid sequence <SEQ ID 4296>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1891(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 546/1075 (50%), Positives = 758/1075 (69%), Gaps = 11/1075 (1%)

Query: 1    MKLLYTDINHDMTEILVNQAAHAAEAGWRIFYIAPNSLSFEKERAVLENLPQEASFAITI 60
            MKL+YT++++ MTEILVN+A  AA+ G+R+FYIAPNSLSFEKER VL  LP+  +F+I +
Sbjct: 1    MKLIYTEMSYSMTEILVNEARKAADQGYRVFYIAPNSLSFEKEREVLTLLPERGTFSIIV 60

Query: 61   TRFAQLARYFTLNQPNQKESLNDIGLAMIFYRALASFEDGQLKVFGRLKQDASFISQLVD 120
            TRF Q++RYFT+     K+ L+D  LAMIFYRAL   +   L  +GRL+ ++ FI QLV+
Sbjct: 61   TRFVQMSRYFTVESSPSKQHLDDTTLAMIFYRALMQLKPEDLPSYGRLQNNSVFIEQLVE 120

Query: 121  LYKELQTANLSILELKYLHSPEKFEDLLAIFLVVSDLLREGEYDNQSKIAFFTEQVRSGQ 180
            LYKEL+ A LS+ +L  L  P+K EDL+ I  +   ++ + +Y+  S + F   ++ G
Sbjct: 121  LYKELKNAQLSVHDLTGLDHPQKQEDLIKIIELAETIMIQQDYNQDSPLQSFARAIKLGL 180

Query: 181  LDVDLKNTILIVDGFTRFSAEEEALIKSLSSRCQEIIIGAYASQKAYKANFTNGNIYSAG 240
            L+  L  T++++DGF+RFSAEE+ L+  L++ CQE+IIG+Y SQKAY+ +F  GNIY A
Sbjct: 181  LNNQLSKTVVVIDGFSRFSAEEDYLLSLLNNNCQEVIIGSYVSQKAYQKSFIKGNIYEAS 240

Query: 241  VDFLRYLATTFQTKPEFILSKWESKSGFEMISKNIEGKHDFTNSSHILDDTAKDCITIWE 300
            + FL+ LA  +  KP F  S    K  F  +++  E  HDF+     L  +  D  ++W+
Sbjct: 241  LHFLQDLAQKYHIKPVFATSNQVFKPAFSRLTQLFEATHDFSQVDWQLQKSDLDHFSLWQ 300

Query: 301  CINQKDEVEHVARAIRQKLYQGYRYKDILVLLGDVDSYKLQLSKIFEQYDIPYYFGKAET 360
            C +QK+E+EHVA++IRQKLY+GYRYKDILVLLGD+D+Y+LQ+  IF++++IPYY GKAE
Sbjct: 301  CHHQKEEIEHVAKSIRQKLYEGYRYKDILVLLGDMDAYQLQIGPIFDKFEIPYYLGKAEP 360

Query: 361  MAAHPLVHFMDSLSRIKRYRFRAEDVLNLFKTGIYGEISQDDLDYFEAYISYADIKGPKK 420
            MAAHPLV F++SL R +RY +R ED+LN+ K+G++G     D+D FE Y  +ADIKG  K
Sbjct: 361  MAAHPLVQFIESLERSQRYNWRREDILNMLKSGLFGCFDDSDIDRFEEYTQFADIKGFTK 420

Query: 421  FFTDFVV-GAKKFDLGRLNTIRQSLLTPLESFVKTKKQDGIKTLNQFMFFLTQVGLSDNL 479
            F   F +  ++++ L  LN +RQ ++ PL+   K++KQ G   +++ + FL ++ L++N+
Sbjct: 421  FSKPFTINSSRQYPLDFLNEMRQDIVLPLQELFKSQKQLGASLVDKLILFLKKIRLAENM 480

Query: 480  SRLVGQMSENEQEKHQEVWKTFTDILEQFQTIFGQEKLNLDEFLSLLNSGMMQAEYRMVP 539
              L    S+ E EK++EVWK FTDIL  F  IFGQEKL L + L+L+ +GM  A+YR+VP
Sbjct: 481  QGLA--QSQLEVEKNEEVWKRFTDILTSFHHIFGQEKLRLSDCLALIKTGMKSAQYRVVP 538

Query: 540  ATVDVVTVKSYDLVEPHSNQFVYALGMTQSHFPKIAQNKSLISDIERQLINDANDTDGHF 599
            AT+DVVT+KSYDLV+PHS  FVYA+G+TQSHFPK   +  L+SD ER  IN+  +   HF
Sbjct: 539  ATLDVVTIKSYDLVQPHSKPFVYAIGLTQSHFPKQIHHSGLLSDQERARINEIRNY-RHF 597
```

```
Query: 600  DIMTQENLKKNHFAALSLFNAAKQELVLTIPQLLNESEDQMSPYLVELRDIGVPFNHKGR 659
            DI + EN KKNH  ALSLFNAA +ELVL++  ++NE+ D +SPYL EL + G+P   KG+
Sbjct: 598  DIASAENSKKNHQTALSLFNAATKELVLSVSTVINETFDDLSPYLKELINFGLPLLDKGK 657

Query: 660  QSLKEEADNIGNYKALLSRVVDLYRSAIDKEMTKEEQTFWSVAVRYLRRQLTSKGIEIPI 719
              L  +  +IGNYKALLS+++ + R  +  EM+ +++ FW+V +RYLR+QL  + +E+P
Sbjct: 658  NYLSYDNSDIGNYKALLSQIIAINRQDL-IEMSDQDKMFWTVVLRYLRKQLRKQQLELPT 716

Query: 720  ITDSLDTVTVSSDVMTRRFPEDDPLKLSSSALTTFYNNQYKYFLQYVLGLEEQDSIHPDM 779
                L T  +S +V+   FP+  PLKLS++ALT FYNNQY YFL+YVL L + +SIHPD
Sbjct: 717  SDYRLSTKPLSKEVIEVCFPKGIPLKLSATALTVFYNNQYNYFLKYVLNLNKTESIHPDS 776

Query: 780  RHHGTYLHRVFEILMKNQGIESFEEKLNSAINKTNQEDVFKSLYSEDAESRYSLEILEDI 839
            R HG YLHRVFE LMK+   E F+ KL  AI  TNQE  F+ +Y ++AE+ YSL ILEDI
Sbjct: 777  RIHGQYLHRVFERLMKDHTQEPFDNKLKQAIYHTNQESFFQQVYQDNAEAEYSLAILEDI 836

Query: 840  ARATATILRQDSQMTVESEEERFELMIDNTIKINGIIDRIDRLSDGSLGVVDYKSSAQKF 899
             R+TA IL+ +  + V  +E+ F+L + N I ++GIIDRID+LSDGSLG+VDYKSSA +F
Sbjct: 837  VRSTAPILQLNQNIQVIDQEKNFQLDMGNEILVHGIIDRIDQLSDGSLGIVDYKSSANQF 896

Query: 900  DIQKFYNGLSPQLVTYIDAISR--DKEVEQKPPIFGAMYLHMQEPRQDLSKIKNLDD-LV 956
            DI  FYNGLSPQL+TY+ A+ +    ++ Q   +FGAMYLH+Q+P+ DL   K +D+ LV
Sbjct: 897  DIGTFYNGLSPQLMTYLAALKQIAPHDINQ---LFGAMYLHLQDPKLDLVTFKQIDNTLV 953

Query: 957  TKNHQALTYKGLFSEAEKEFLANGKYHLKDSLYSETEIAILQAHNQSLYKKASETIKSGK 1016
               ++ALTYKG+FSE EKE L+ G Y  K++LYS  E+  L  +N+ LY KA++ IK G
Sbjct: 954  ESIYKALTYKGIFSEVEKEHLSTGAYQTKNALYSNDELETLLNYNKYLYLKAAKHIKKGH 1013

Query: 1017 FLINPYTEDAKTVDGDQFKSITGFEADRHMARARALYKLPAKEKRQGFLTLMQQE 1071
            FLINPYT D KTV GDQ K+IT FEAD  M +AR L  LPAKEK++ FLTLM++E
Sbjct: 1014 FLINPYTSDGKTVQGDQLKAITRFEADLDMGQARRLVTLPAKEKKECFLTLMRKE 1068
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1400

A DNA sequence (GBSx1485) was identified in *S.agalactiae* <SEQ ID 4297> which encodes the amino acid sequence <SEQ ID 4298>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.80   Transmembrane   51 -  67 (  44 -  69)

----- Final Results -----
              bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8799> which encodes amino acid sequence <SEQ ID 8800> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -20.62
GvH: Signal Score (-7.5): -6.25
     Possible site: 31
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -7.80 threshold:  0.0
   INTEGRAL    Likelihood = -7.80   Transmembrane   47 -  63 (  40 -  65)
   PERIPHERAL  Likelihood =  3.34      26
 modified ALOM score:   2.06

*** Reasoning Step: 3
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC75528 GB:AE000334 orf, hypothetical protein [Escherichia coli K12]
 Identities = 138/297 (46%), Positives = 193/297 (64%), Gaps = 16/297 (5%)

Query: 5   MKIDDLRKSDNVEDRRSSSGGSFSSGGSGLPILQLLLLRGSWKTKLVVLIILLLLG--GG 62
           M+    R+SDNVEDRR+SSGG  S GG G  +        S K  L++LI++L+ G  G
Sbjct: 1   MRWQGRRESDNVEDRRNSSGGP-SMGGPGFRL-------PSGKGGLILLIVVLVAGYYGV 52

Query: 63  GLTSIFNDSSSPSSYQSQNVSRSVDNSATREQIDFVNKVLGSTEDFWSQEFQTQGFGNYK 122
            LT +           ++++S + D +A      F + +L +TED W Q+F+  G   Y+
Sbjct: 53  DLTGLMTGQPVSQQQSTRSISPNEDEAAK-----FTSVILATTEDTWGQQFEKMG-KTYQ 106

Query: 123 EPKLVLYTNSIQTGCGIGESASGPFYCSADKKIYLDISFYNELSHKYGATGDFAMAYVIA 182
           +PKLV+Y    +TGCG G+S  GPFYC AD  +Y+D+SFY+++  K GA GDFA  YVIA
Sbjct: 107 QPKLVMYRGMTRTGCGAGQSIMGPFYCPADGTVYIDLSFYDDMKDKLGADGDFAQGYVIA 166

Query: 183 HEVGHHIQTELGIMDKYNRMRHGLTKKEANALNVRLELQADYYAGVWAHYIRGKNLLEQG 242
           HEVGHH+Q  LGI  K  +++   T+ E N L+VR+ELQAD +AGVW H ++ + +LE G
Sbjct: 167 HEVGHHVQKLLGIEPKVRQLQQNATQAEVNRLSVRMELQADCFAGVWGHSMQQQGVLETG 226

Query: 243 DFEEAMNAAHAVGDDTLQKETYGKLVPDSFTHGTAEQRQRWFNKGFQYGDIQHGDTF 299
           D EEA+NAA A+GDD LQ+++ G++VPDSFTHGT++QR  WF +GF  GD    +TF
Sbjct: 227 DLEEALNAAQAIGDDRLQQQSQGRVVPDSFTHGTSQQRYSWFKRGFDSGDPAQCNTF 283
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4299> which encodes the amino acid sequence <SEQ ID 4300>. Analysis of this protein sequence reveals the following:

```
      Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.42  Transmembrane   48 -  64 (  41 -  67)

----- Final Results -----
               bacterial membrane --- Certainty=0.3569(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC75528 GB:AE000334 orf, hypothetical protein [Escherichia coli]
 Identities = 143/301 (47%), Positives = 195/301 (64%), Gaps = 21/301 (6%)

Query: 1   MKTDDLRESQQVEDRRGQSSG-SFGGGGLGGGLLLQLLFSRGGWKTKLVILLLLLVMG-- 57
           M+    RES  VEDRR  S G S GG G       +L   +GG    L++L+++LV G
Sbjct: 1   MRWQGRRESDNVEDRRNSSGGPSMGGPGF------RLPSGKGG----LILLIVVLVAGYY 50

Query: 58  GGGLSGVLGGKPSSTNNNAYQSSQVTRTNGDKASQEQVSFVSKVFASTEDYWTKTFREKG 117
           G  L+G++ G+P S      QS++    N D+A++    F S + A+TED W + F + G
Sbjct: 51  GVDLTGLMTGQPVSQQ----QSTRSISPNEDEAAK----FTSVILATTEDTWGQQFEKMG 102

Query: 118 LTYHKPTLVLYTGATQTACGRGQASSGPFYCPGDQKVYLDISFYNELSTKYGAKGDFAMA 177
            TY +P LV+Y G T+T CG GQ+  GPFYCP D  VY+D+SFY+++  K GA GDFA
Sbjct: 103 KTYQQPKLVMYRGMTRTGCGAGQSIMGPFYCPADGTVYIDLSFYDDMKDKLGADGDFAQG 162

Query: 178 YVIAHEVGHHIQNELGIMDNYASARQGKSKAKANQLNVKLELQADYYAGAWANYVQGQGL 237
           YVIAHEVGHH+Q  LGI       +Q  ++A+ N+L+V++ELQAD +AG W + +Q QG+
Sbjct: 163 YVIAHEVGHHVQKLLGIEPKVRQLQQNATQAEVNRLSVRMELQADCFAGVWGHSMQQQGV 222

Query: 238 LEKGDIEEAMAAAHAVGDDTLQEETYGRTVPDSFTHGTSKQRQRWFDRGYQYGDFEHGDTF 298
           LE GD+EEA+ AA A+GDD LQ+++ GR VPDSFTHGTS+QR  WF RG+  GD    +TF
Sbjct: 223 LETGDLEEALNAAQAIGDDRLQQQSQGRVVPDSFTHGTSQQRYSWFKRGFDSGDPAQCNTF 283
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 191/303 (63%), Positives = 241/303 (79%), Gaps = 5/303 (1%)

Query: 5   MKIDDLRKSDNVEDRRSSSGGSFSSGG-SGLPILQLLLLRGSWKTKLVVLIILLLLGGGG 63
           MK DDLR+S  VEDRR  S GSF  GG  G  +LQLL  RG WKTKLV+L++LL++GGGG
Sbjct: 1   MKTDDLRESQQVEDRRGQSSGSFGGGGLGGGLLLQLLFSRGGWKTKLVILLLLLVMGGGG 60

Query: 64  LTSIFN---DSSSPSSYQSQNVSRSVDNSATREQIDFVNKVLGSTEDFWSQEFQTQGFGN 120
           L+ +      S++ ++YQS  V+R+  + A++EQ+ FV+KV  STED+W++ F+ +G
Sbjct: 61  LSGVLGGKPSSTNNNAYQSSQVTRTNGDKASQEQVSFVSKVFASTEDYWTKTFREKGL-T 119

Query: 121 YKEPKLVLYTNSIQTGCGIGESASGPFYCSADKKIYLDISFYNELSHKYGATGDFAMAYV 180
           Y +P LVLYT + QT CG G+++SGPFYC  D+K+YLDISFYNELS KYGA GDFAMAYV
Sbjct: 120 YHKPTLVLYTGATQTACGRGQASSGPFYCPGDQKVYLDISFYNELSTKYGAKGDFAMAYV 179

Query: 181 IAHEVGHHIQTELGIMDKYNRMRHGLTKKEANALNVRLELQADYYAGVWAHYIRGKNLLE 240
           IAHEVGHHIQ ELGIMD Y   R G +K +AN LNV+LELQADYYAG WA+Y++G+ LLE
Sbjct: 180 IAHEVGHHIQNELGIMDNYASARQGKSKAKANQLNVKLELQADYYAGAWANYVQGQGLLE 239

Query: 241 QGDFEEAMNAAHAVGDDTLQKETYGKLVPDSFTHGTAEQRQRWFNKGFQYGDIQHGDTFS 300
           +GD EEAM AAHAVGDDTLQ+ETYG+ VPDSFTHGT++QRQRWF++G+QYGD +HGDTFS
Sbjct: 240 KGDIEEAMAAAHAVGDDTLQEETYGRTVPDSFTHGTSKQRQRWFDRGYQYGDFEHGDTFS 299

Query: 301 VEH 303
           + +
Sbjct: 300 IPY 302
```

SEQ ID 8800 (GBS404) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 3; MW 62kDa).

GBS404-GST was purified as shown in Figure 218, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1401

A DNA sequence (GBSx1486) was identified in *S.agalactiae* <SEQ ID 4301> which encodes the amino acid sequence <SEQ ID 4302>. This protein is predicted to be phenylalanyl-tRNA synthetase beta chain (pheT). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2617(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14823 GB:Z99118 phenylalanyl-tRNA synthetase (beta subunit)
          [Bacillus subtilis]
 Identities = 376/805 (46%), Positives = 523/805 (64%), Gaps = 6/805 (0%)

Query: 1   MLVSYKWLKELVDVD-VTTAELAEKMSTTGIEVEGVETPAEGLSKLVVGHIVSCEDVPDT 59
           M VSYKWL++ VD+  +  A LAEK++  GIEVEG+E   EG+  +V+GH++  E  P+
Sbjct: 1   MFVSYKWLEDYVDLKGMDPAVLAEKITRAGIEVEGIEYKGEGIKGVVIGHVLEREQHPNA 60

Query: 60  H-LHLCQVDTGDDELRQVVCGAPNVKTGINVIVAVPGARIADNYKIKKGKIRGMESLGMI 118
             L+ C VD G +    Q++CGAPNV  G  V VA  GA +  N+KIKK K+RG ES GMI
Sbjct: 61  DKLNKCLVDIGAEAPVQIICGAPNVDKGQKVAVATVGAVLPGNFKIKKAKLRGEESNGMI 120
```

```
Query: 119 CSLQELGLSESIIPKEFSDGIQILPEGAIPGDSIFSYLDLDDEIIELSITPNRADALSMR 178
           CSLQELG+   ++ KE+++GI + P  A  G    + L LDD I+EL +TPNRADA++M
Sbjct: 121 CSLQELGIESKLVAKEYAEGIFVFPNDAETGSDALAALQLDDAILELGLTPNRADAMNML 180

Query: 179 GVAHEVAAIYGKKVHFEEKNLIEEAERAADKISVVIESDKVLS-YSARIVKNVTVAPSPQ 237
           GVA+EVAAI   +V   + +    +E+A+D ISV IE  +    Y+A+I+KNVT+APSP
Sbjct: 181 GVAYEVAAILDTEVKLPQTDYPAASEQASDYISVKIEDQEANPLYTAKIIKNVTIAPSPL 240

Query: 238 WLQNKLMNAGIRPINNVVDVTNYVLLTYGQPMHAFDFDKFDGTTIVARNAENGEKLITLD 297
           W+Q KLMNAGIRP NNVVD+TN+VLL YGQP+HAFD+D+F    +V R A   E ++TLD
Sbjct: 241 WMQTKLMNAGIRPHNNVVDITNFVLLEYGQPLHAFDYDRFGSKEVVVRKAAENEMIVTLD 300

Query: 298 GEERDLIADDLVIAVNDQPVALAGVMGGQSTEIGSSSKTVVLEAAVFNGTSIRKTSGRLN 357
            +ER L AD LVI    +  A+AGVMGG  +E+   +KT++LEAA FNG  +RK S  L
Sbjct: 301 DQERKLSADHLVITNGTKAQAVAGVMGGAESEVQEDTKTILLEAAYFNGQKVRKASKDLG 360

Query: 358 LRSESSSRFEKGINYDTVSEAMDFAAAMLQELAGGQVLSGQVTEGVLPTEPVEVSTTLGY 417
           LRSESS RFEKGI+   V  A + AA ++   AGG+VL+G V E  L  E   +  +
Sbjct: 361 LRSESSVRFEKGIDPARVRLAAERAAQLIHLYAGGEVLAGTVEEDHLTIEANNIHVSADK 420

Query: 418 VNTRLGTELTYTDIEEVFEKLGFAISGSEVKFTVLVPRRRWDIAIQADLVEEIARIYGYE 477
           V++ LG  ++  ++  ++++LGF +  ++    V VP RR DI I+ DL+EE AR+YGY+
Sbjct: 421 VSSVLGLTISKEELISIYKRLGFTVGEADDLLVVTVPSRRGDITIEEDLIEEAARLYGYD 480

Query: 478 KLPTTLPEAGATAGELTSMQRLRRRVRTVAEGAGLSEIITYALTTPEKAVQFSTQATNIT 537
            +P+TLPE   T G LT  Q  RR+VR   EGAGLS+ ITY+LT  +KA  F+ + +  T
Sbjct: 481 NIPSTLPETAGTTGGLTPYQAKRRKVRRFLEGAGLSQAITYSLTNEKKATAFAIEKSLNT 540

Query: 538 ELMWPMTVDRSALRQNVVSGMLDTIAYNVARKNSNLAVYEIGKVFEQTGNPKEDLPTEVE 597
            L  PM+ +RS LR ++V  +LD+++YN+AR+  ++A+YE+G VF      ++  P E E
Sbjct: 541 VLALPMSEERSILRHSLVPNLLDSVSYNLARQTDSVALYEVGSVF--LTKEEDTKPVETE 598

Query: 598 TFTFALTGLVEEKDFQTKSKPVDFFYAKGIVEALFIKLK-LDVTFVAQKGLASMHPGRTA 656
               A+TGL  ++ +Q + KPVDFF  KGIVE L  KL  LD     Q     +HPGRTA
Sbjct: 599 RVAGAVTGLWRKQLWQGEKKPVDFFVVKGIVEGLLDKLNVLDSIEFVQSERKQLHPGRTA 658

Query: 657 TILLDGKEIGFVGQVHPQTAKQYDIPETYVAEINLSTIESQMNQALIFEDITKYPSVSRD 716
            ILL+G  IGF+GQVHP   K+ DI ETYV E++L  + +     L++  I KYPSV+RD
Sbjct: 659 NILLNGSLIGFIGQVHPSLEKELDIKETYVFELDLHALLAAETAPLVYTAIPKYPSVTRD 718

Query: 717 IALLLAESVSHHDIVSAIETSGVKRLTAIKLFDVYAGNNIAEGYKSMAYSLTFQNPNDNL 776
           IAL+  ++V+   + S I+ +G K L  + +FDVY G ++ EG KS+A+SL + NP   L
Sbjct: 719 IALVTDKTVTSGQLESVIKEAGGKLLKEVTVFDVYEGEHMEEGKKSVAFSLQYVNPEQTL 778

Query: 777 TDEEVAKYMEKITKSLVEKVNAEIR 801
           T+EEV K   K+ K+L +   A +R
Sbjct: 779 TEEEVTKAHSKVLKALEDTYQAVLR 803
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4303> which encodes the amino acid sequence <SEQ ID 4304>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1283(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 595/801 (74%), Positives = 687/801 (85%)

Query: 1   MLVSYKWLKELVDVDVTTAELAEKMSTTGIEVEGVETPAEGLSKLVVGHIVSCEDVPDTH 60
           MLVSYKWLKELVD+DVT A LAEKMSTTGIEVEG+E PA+GLSKLVVGH++SCEDVP+TH
Sbjct: 6   MLVSYKWLKELVDIDVTPAALAEKMSTTGIEVEGIEVPADGLSKLVVGHVLSCEDVPETH 65
```

```
Query: 61  LHLCQVDTGDDELRQVVCGAPNVKTGINVIVAVPGARIADNYKIKKGKIRGMESLGMICS 120
           LHLCQVDTGD+  RQ+VCGAPNVK GI VIVAVPGARIADNYKIKKGKIRGMESLGMICS
Sbjct: 66  LHLCQVDTGDETPRQIVCGAPNVKAGIKVIVAVPGARIADNYKIKKGKIRGMESLGMICS 125

Query: 121 LQELGLSESIIPKEFSDGIQILPEGAIPGDSIFSYLDLDDEIIELSITPNRADALSMRGV 180
           LQELGLS+SIIPKEFSDGIQILPE A+PGD+IF YLDLDD IIELSITPNRADALSMRGV
Sbjct: 126 LQELGLSDSIIPKEFSDGIQILPEEAVPGDAIFKYLDLDDHIIELSITPNRADALSMRGV 185

Query: 181 AHEVAAIYGKKVHFEEKNLIEEAERAADKISVVIESDKVLSYSARIVKNVTVAPSPQWLQ 240
           AHEVAAIYGK V F +KNL E  +  ++ I V I SD VL+Y++R+V+NV V PSPQWLQ
Sbjct: 186 AHEVAAIYGKSVSFPQKNLQESDKATSEAIEVAIASDNVLTYASRVVENVKVKPSPQWLQ 245

Query: 241 NKLMNAGIRPINNVVDVTNYVLLTYGQPMHAFDFDKFDGTTIVARNAENGEKLITLDGEE 300
           N LMNAGIRPINNVVDVTNYVLL +GQPMHAFD+DKF+   IVAR A  GE L+TLDGE+
Sbjct: 246 NLLMNAGIRPINNVVDVTNYVLLYFGQPMHAFDYDKFEDHKIVARAARQGESLVTLDGEK 305

Query: 301 RDLIADDLVIAVNDQPVALAGVMGGQSTEIGSSSKTVVLEAAVFNGTSIRKTSGRLNLRS 360
           RDL  +DLVI V D+PVALAGVMGGQ+TEI ++S+TVVLEAAVF+G SIRKTSGRLNLRS
Sbjct: 306 RDLTTEDLVITVADKPVALAGVMGGQATEIDANSQTVVLEAAVFDGKSIRKTSGRLNLRS 365

Query: 361 ESSSRFEKGINYDTVSEAMDFAAAMLQELAGGQVLSGQVTEGVLPTEPVEVSTTLGYVNT 420
           ESSSRFEKG+NY TV EA+DFAAAMLQELA GQVLSG V  G LPTEPVEVST+L YVN
Sbjct: 366 ESSSRFEKGVNYATVLEALDFAAAMLQELAEGQVLSGHVQAGQLPTEPVEVSTSLDYVNV 425

Query: 421 RLGTELTYTDIEEVFEKLGFAISGSEVKFTVLVPRRRWDIAIQADLVEEIARIYGYEKLP 480
           RLGTELT+ DI+ +F++LGF ++G E  FTV VPRRRWD++I ADLVEEIARIYGY+KLP
Sbjct: 426 RLGTELTFADIQRIFDQLGFGLTGDETSFTVAVPRRRWDVSIPADLVEEIARIYGYDKLP 485

Query: 481 TTLPEAGATAGELTSMQRLRRRVRTVAEGAGLSEIITYALTTPEKAVQFSTQATNITELM 540
           TTLPEAG TA ELT  Q LRR+VR +AEG GL+EII+YALTTPEKAV+F+   +++TELM
Sbjct: 486 TTLPEAGGTAAELTPTQALRRKVRGLAEGLGLTEIISYALTTPEKAVEFAVAPSHLTELM 545

Query: 541 WPMTVDRSALRQNVVSGMLDTIAYNVARKNSNLAVYEIGKVFEQTGNPKEDLPTEVETFT 600
           WPM+V+RSALRQN+VSGMLDT+AYNVARK SNLA+YEIGK+FEQ  NPKEDLP EV  F
Sbjct: 546 WPMSVERSALRQNMVSGMLDTVAYNVARKQSNLALYEIGKIFEQEANPKEDLPNEVNHFA 605

Query: 601 FALTGLVEEKDFQTKSKPVDFFYAKGIVEALFIKLKLDVTFVAQKGLASMHPGRTATILL 660
           FA+ GLV +KDFQT+++ VDF++AKG ++ LF  L L V +V  K LA+MHPGRTA ILL
Sbjct: 606 FAICGLVAQKDFQTQAQAVDFYHAKGNLDTLFANLNLKVQYVPTKDLANMHPGRTALILL 665

Query: 661 DGKEIGFVGQVHPQTAKQYDIPETYVAEINLSTIESQMNQALIFEDITKYPSVSRDIALL 720
           D + IGFVGQVHP TAK Y IPETYVAE++++ +E+ +     F +ITK+P+++RD+ALL
Sbjct: 666 DEQVIGFVGQVHPGTAKAYSIPETYVAELDMAALEAALPSDQTFAEITKFPAMTRDVALL 725

Query: 721 LAESVSHHDIVSAIETSGVKRLTAIKLFDVYAGNNIAEGYKSMAYSLTFQNPNDNLTDEE 780
           L   VSH  IV+AIE++GVKRLT+IKLFDVY G  I  G KSMAYSLTFQNPNDNLTDEE
Sbjct: 726 LDREVSHQAIVTAIESAGVKRLTSIKLFDVYEGATIQAGKKSMAYSLTFQNPNDNLTDEE 785

Query: 781 VAKYMEKITKSLVEKVNAEIR 801
           VAKYMEKITK+L E+V AE+R
Sbjct: 786 VAKYMEKITKALTEQVGAEVR 806
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1402

A DNA sequence (GBSx1487) was identified in *S.agalactiae* <SEQ ID 4305> which encodes the amino acid sequence <SEQ ID 4306>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0653(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9769> which encodes amino acid sequence <SEQ ID 9770> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15205 GB:Z99120 transcriptional regulator [Bacillus subtilis]
 Identities = 60/169 (35%), Positives = 100/169 (58%)

Query: 17  ITFKKVGLDNVNILQNIAIETFRQTFSHDNSEEQLQAFFNESYTLPVLKSEITHAESDTY 76
           +  KK   +++  LQ ++IETF  TF   NS E ++A+   ++    L+ E+++  S  +
Sbjct: 3   VKMKKCSREDLQTLQQLSIETFNDTFKEQNSPENMKAYLESAFNTEQLEKELSNMSSQFF 62

Query: 77  FVYLDTDLVGYLKVNWGSQQTEKDLDKAFEIQRIYLLDAYQGQGIGKATFEFALDLAYKS 136
           F+Y D ++ GY+KVN    Q+E+   ++ EI+RIY+ +++Q  G+GK     A+++A +
Sbjct: 63  FIYFDHEIAGYVKVNIDDAQSEEMGAESLEIERIYIKNSFQKHGLGKHLLNKAIEIALER 122

Query: 137 GLDWAWLGVWEFNHKAQAFYAKYGFEKFSEHQFSVGDKVDTDWLLRKSL 185
                WLGVWE N  A AFY K GF +   H F +GD+  TD ++ K+L
Sbjct: 123 NKKNIWLGVWEKNENAIAFYKKMGFVQTGAHSFYMGDEEQTDLIMAKTL 171
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1403

A DNA sequence (GBSx1488) was identified in *S.agalactiae* <SEQ ID 4307> which encodes the amino acid sequence <SEQ ID 4308>. This protein is predicted to be phenylalanyl-tRNA synthetase (alpha subunit) (pheS). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3937(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9339> which encodes amino acid sequence <SEQ ID 9340> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14824 GB:Z99118 phenylalanyl-tRNA synthetase (alpha subunit)
          [Bacillus subtilis]
 Identities = 209/338 (61%), Positives = 270/338 (79%), Gaps = 2/338 (0%)

Query: 1   MKISTQEKLKEM-TGNHTKELQDLRVQVLGKKGSLTELLKGLKDLSNDLRPVVGKQVNEV 59
           +K   QE L+++   +  K + D+RVQ LGKKG +TE+L+G+  LS + RP +G   NEV
Sbjct: 5   LKQLEQEALEQVEAASSLKVVNDIRVQYLGKKGPITEVLRGMGKLSAEERPKMGALANEV 64

Query: 60  RDILTKAFEEQAKVVEAAKIQAQLESESVDVTLPGRQMTLGHRHVLTQTSEEIEDIFLGM 119
           R+ +  A  ++ + +E  +++ +L  +++DVTLPG  + +G RH LT   EEIED+F+GM
Sbjct: 65  RERIANAIADKNEKLEEEEMKQKLAGQTIDVTLPGNPVAVGGRHPLTVVIEEIEDLFIGM 124

Query: 120 GFQVVDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHTSPVQARTMDQHDF 179
           G+ V +G EVE DYYNFE +NLPK+HPARDMQD+FYITEE L+RT TSPVQ RTM++H+
Sbjct: 125 GYTVEEGPEVETDYYNFESLNLPKEHPARDMQDSFYITEETLMRTQTSPVQTRTMEKHE- 183

Query: 180 SKGPLKMISPGRVFRRDTDDATHSHQFHQIEGLVVGENISMGDLKGTLQLISQKMFGAER 239
```

```
           KGP+K+I PG+V+RRD DDATHSHQF QIEGLVV +NISM DLKGTL+L+++KMFG +R
Sbjct: 184 GKGPVKIICPGKVYRRDNDDATHSHQFMQIEGLVVDKNISMSDLKGTLELVAKKMFGQDR 243

Query: 240 KIRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKQTGWIEILGAGMVHPSVLEMSGIDSE 299
           +IRLRPS+FPFTEPSVEVDV+CFKCGG GC+VCK TGWIEILGAGMVHP+VL+M+G D +
Sbjct: 244 EIRLRPSFFPFTEPSVEVDVTCFKCGGNGCSVCKGTGWIEILGAGMVHPNVLKMAGFDPK 303

Query: 300 KYSGFAFGLGQERIAMLRYGINDIRGFYQGDVRFTDQF 337
           +Y GFAFG+G ERIAML+YGI+DIR FY  DVRF  QF
Sbjct: 304 EYQGFAFGMGVERIAMLKYGIDDIRHFYTNDVRFISQF 341
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4309> which encodes the amino acid sequence <SEQ ID 4310>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2806(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 305/337 (90%), Positives = 327/337 (96%)

Query: 1   MKISTQEKLKEMTGNHTKELQDLRVQVLGKKGSLTELLKGLKDLSNDLRPVVGKQVNEVR 60
           +K  T E L+ +TGNHTKELQDLRV VLGKKGSLTELLKGLKDLSNDLRPVVGKQVNEVR
Sbjct: 36  LKTKTLETLQSLTGNHTKELQDLRVAVLGKKGSLTELLKGLKDLSNDLRPVVGKQVNEVR 95

Query: 61  DILTKAFEEQAKVVEAAKIQAQLESESVDVTLPGRQMTLGHRHVLTQTSEEIEDIFLGMG 120
           D+LTKAFEEQAK+VEAAKIQAQL++ES+DVTLPGRQMTLGHRHVLTQTSEEIEDIFLGMG
Sbjct: 96  DLLTKAFEEQAKIVEAAKIQAQLDAESIDVTLPGRQMTLGHRHVLTQTSEEIEDIFLGMG 155

Query: 121 FQVVDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHTSPVQARTMDQHDFS 180
           FQ+VDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHTSPVQART+DQHDFS
Sbjct: 156 FQIVDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHTSPVQARTLDQHDFS 215

Query: 181 KGPLKMISPGRVFRRDTDDATHSHQFHQIEGLVVGENISMGDLKGTLQLISQKMFGAERK 240
           KGPLKM+SPGRVFRRDTDDATHSHQFHQIEGLVVG+NISMGDLKGTL++I +KMFG ER
Sbjct: 216 KGPLKMVSPGRVFRRDTDDATHSHQFHQIEGLVVGKNISMGDLKGTLEMIIKKMFGDERS 275

Query: 241 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKQTGWIEILGAGMVHPSVLEMSGIDSEK 300
           IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCK+TGWIEILGAGMVHPSVLEMSG+D+++
Sbjct: 276 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKKTGWIEILGAGMVHPSVLEMSGVDAKE 335

Query: 301 YSGFAFGLGQERIAMLRYGINDIRGFYQGDVRFTDQF 337
           YSGFAFGLGQERIAMLRYGINDIRGFYQGD RF++QF
Sbjct: 336 YSGFAFGLGQERIAMLRYGINDIRGFYQGDQRFSEQF 372
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1404

A DNA sequence (GBSx1489) was identified in *S.agalactiae* <SEQ ID 4311> which encodes the amino acid sequence <SEQ ID 4312>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2834(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1405

A DNA sequence (GBSx1490) was identified in *S.agalactiae* <SEQ ID 4313> which encodes the amino acid sequence <SEQ ID 4314>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2762(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1406

A DNA sequence (GBSx1491) was identified in *S.agalactiae* <SEQ ID 4315> which encodes the amino acid sequence <SEQ ID 4316>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8801> which encodes amino acid sequence <SEQ ID 8802> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:     10.13
GvH: Signal Score (-7.5): -5.07
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -6.79 threshold:  0.0
   INTEGRAL    Likelihood = -6.79   Transmembrane    8 -  24 (   6 -  27)
   PERIPHERAL  Likelihood =  6.26     258
 modified ALOM score:   1.86

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
 Identities = 154/232 (66%), Positives = 180/232 (77%), Gaps = 1/232 (0%)

Query: 41  KNVSGTPSRELSESVLTSNVKKQLGTNIAWNQSGAFIINQNKTDLNAKVSSAPYAINEIK 100
            K  S  PS+ L+ESVLT  VK Q+  ++ WN SGAFI+N NKT+L+AKVSS PYA N+ K
Sbjct: 43  KQASEAPSQALAESVLTDAVKSQIKGSLEWNGSGAFIVNGNKTNLDAKVSSKPYADNKTK 102

Query: 101 KVNNQIVPTKANALLTKATRQYRNREETGNGRTYWKPAGWHQINGLKGSYNHAVDRGHLI 160
             V  + VPT ANALL+KATRQY+NR+ETGNG T W P GWHQ+  LKGSY HAVDRGHL+
Sbjct: 103 TVGKETVPTVANALLSKATRQYKNRKETGNGSTSWTPPGWHQVKNLKGSYTHAVDRGHLL 162

Query: 161 GYALVGSLRGFDASTSNPKNIATQAAWANQANSNQSTGQNYYETLVRKALDRHKTVRYRV 220
            GYAL+G L GFDASTSNPKNIA Q AWANQA +  STGQNYYE+ VRKALD++K VRYRV
Sbjct: 163 GYALIGGLDGFDASTSNPKNIAVQTAWANQAQAEYSTGQNYYESKVRKALDQNKRVRYRV 222

Query: 221 TLIY-DRDNLLSSGSHIEAKSSDGSLEFNVFIPNVQSGLLFDYATGKVKQTK 271
            TL Y   ++L+ S S IEAKSSDG LEFNV +PNVQ GL  DY TG+V  T+
Sbjct: 223 TLYYASNEDLVPSASQIEAKSSDGELEFNVLVPNVQKGLQLDYRTGEVTVTQ 274
```

There is also homology to SEQ IDs 368 and 1302.

SEQ ID 8802 (GBS285) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 6; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 7; MW 57.5kDa).

GBS285-GST was purified as shown in Figure 208 (lane 7) and Figure 225 (lane 8).

GBS658 was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 134 (lane 8 & 9; MW 27kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1407

A DNA sequence (GBSx1492) was identified in *S.agalactiae* <SEQ ID 4317> which encodes the amino acid sequence <SEQ ID 4318>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1408

A DNA sequence (GBSx1493) was identified in *S.agalactiae* <SEQ ID 4319> which encodes the amino acid sequence <SEQ ID 4320>. This protein is predicted to be UDP-N-acetylglucosamine (murA). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1814(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9767> which encodes amino acid sequence <SEQ ID 9768> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15693 GB:Z99122 UDP-N-acetylglucosamine
           1-carboxyvinyltransferase [Bacillus subtilis]
 Identities = 248/423 (58%), Positives = 323/423 (75%), Gaps = 5/423 (1%)

Query: 5   MDKIIVEGGQTQLQGQVVIEGAKNAVLPLLAATILPSQGKTLLTNVPILSDVFTMNNVVR 64
           M+KIIV GGQ +L G V +EGAKNAVLP++AA++L S+ K+++ +VP LSDV+T+N V+R
Sbjct: 1   MEKIIVRGGQ-KLNGTVKVEGAKNAVLPVIAASLLASEEKSVICDVPTLSDVYTINEVLR 59

Query: 65  GLDIQVDFNCDKKEILVDASGDILDVAPYEFVSQMRASIVVLGPILARNGHAKVSMPGGC 124
            L   V F  +  E+ V+AS  +   AP+E+V +MRAS++V+GP+LAR GHA+V++PGGC
Sbjct: 60  HLGADVHF--ENNEVTVNASYALQTEAPFEYVRKMRASVLVMGPLLARTGHARVALPGGC 117

Query: 125 TIGSRPIDLHLKGLEAMGATITQNGGDITAQAE-KLKGANIYMDFPSVGATQNLMMAATL 183
            IGSRPID HLKG EAMGA I    G I A+ + +L+GA IY+DFPSVGAT+NL+MAA L
Sbjct: 118 AIGSRPIDQHLKGFEAMGAEIKVGNGFIEAEVKGRLQGAKIYLDFPSVGATENLIMAAAL 177

Query: 184 ASGTTTIENAAREPEIVDLAQLLNKMGAKVKGAGTETLTIIGVDALHGTEHDVVQDRIEA 243
           A GTTT+EN A+EPEIVDLA  +N MG K++GAGT T+ I GV+ LHG +H ++ DRIEA
Sbjct: 178 AEGTTTLENVAKEPEIVDLANYINGMGGKIRGAGTGTIKIEGVEKLHGVKHHIIPDRIEA 237

Query: 244 GTFMVAAAMTSGNVLVKDAIWEHNRPLISKLMEMGVEVSEEEDGIRVKADTKKLKPVTVK 303
           GTFMVAAA+T GNVLVK A+ EH   LI+K+ EMGV + +E +G+RV    K+LKP+ +K
Sbjct: 238 GTFMVAAAITEGNVLVKGAVPEHLTSLIAKMEEMGVTIKDEGEGLRV-IGPKELKPIDIK 296

Query: 304 TLPHPGFPTDMQAQFTALMAVVNGESTMIETVFENRFQHLEEMRRMGLQTEILRDTAMIH 363
           T+PHPGFPTDMQ+Q  AL+   +G S + ETVFENRF H EE RRM    +I   + +I+
Sbjct: 297 TMPHPGFPTDMQSQMMALLLRASGTSMITETVFENRFMHAEEFRRMNGDIKIEGRSVIIN 356

Query: 364 GGRALQGAPVMSTDLRASAALILAGMVAQGQTVVGQLTHLDRGYYQFHEKLAALGANIKR 423
           G   LQGA V +TDLRA AALILAG+VA+G T V +L HLDRGY  FH+KLAALGA+I+R
Sbjct: 357 GPVQLQGAEVAATDLRAGAALILAGLVAEGHTRVTELKHLDRGYVDFHQKLAALGADIER 416

Query: 424 VSE 426
           V++
Sbjct: 417 VND 419
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4321> which encodes the amino acid sequence <SEQ ID 4322>. Analysis of this protein sequence reveals the following:

```
     Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.03   Transmembrane  377 - 393 ( 376 - 394)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15693 GB:Z99122 UDP-N-acetylglucosamine
           1-carboxyvinyltransferase [Bacillus subtilis]
 Identities = 248/423 (58%), Positives = 318/423 (74%), Gaps = 5/423 (1%)
```

```
Query: 1   VDKIIIEGGQTRLEGEVVIEGAKNAVLPLLAASILPSKGKTILRNVPILSDVFTMNNVVR 60
           ++KII+ GGQ +L G V +EGAKNAVLP++AAS+L S+ K+++ +VP LSDV+T+N V+R
Sbjct: 1   MEKIIVRGGQ-KLNGTVKVEGAKNAVLPVIAASLLASEEKSVICDVPTLSDVYTINEVLR 59

Query: 61  GLDIRVDFNEAANEITVDASGHILDEAPYEYVSQMRASIVVLGPILARNGHAKVSMPGGC 120
            L   V F    NE+TV+AS  +  EAP+EYV +MRAS++V+GP+LAR GHA+V++PGGC
Sbjct: 60  HLGADVHFEN--NEVTVNASYALQTEAPFEYVRKMRASVLVMGPLLARTGHARVALPGGC 117

Query: 121 TIGSRPINLHLKGLEAMGATITQKGGDITAQAD-RLQGAMIYMDFPSVGATQNLMMAATL 179
            IGSRPI+ HLKG EAMGA I    G I A+   RLQGA IY+DFPSVGAT+NL+MAA L
Sbjct: 118 AIGSRPIDQHLKGFEAMGAEIKVGNGFIEAEVKGRLQGAKIYLDFPSVGATENLIMAAAL 177

Query: 180 ADGVTTIENAAREPEIVDLAQFLNKMGARIRGAGTETLTITGVTHLRGVEHDVVQDRIEA 239
           A+G TT+EN A+EPEIVDLA ++N MG +IRGAGT T+ I GV  L GV+H ++ DRIEA
Sbjct: 178 AEGTTTLENVAKEPEIVDLANYINGMGGKIRGAGTGTIKIEGVEKLHGVKHHIIPDRIEA 237

Query: 240 GTFMVAAAMTSGNVLIRDAVWEHNRPLISKLMEMGVSVTEEEYGIRVQANTPKLKPVTVK 299
           GTFMVAAA+T GNVL++ AV EH   LI+K+ EMGV++ +E  G+RV     +LKP+ +K
Sbjct: 238 GTFMVAAAITEGNVLVKGAVPEHLTSLIAKMEEMGVTIKDEGEGLRV-IGPKELKPIDIK 296

Query: 300 TLPHPGFPTDMQAQFTALMAVVNGESTMVETVFENRFQHLEEMRRMGLQSEILRETAMIH 359
           T+PHPGFPTDMQ+Q  AL+   +G S + ETVFENRF H EE RRM    +I   + +I+
Sbjct: 297 TMPHPGFPTDMQSQMMALLLRASGTSMITETVFENRFMHAEEFRRMNGDIKIEGRSVIIN 356

Query: 360 GGRQLQGAPVMSTDLRASAALILTGIVAQGVTIVNNLVHLDRGYYQFHEKLAKLGATISR 419
           G  QLQGA V +TDLRA AALIL G+VA+G T V  L HLDRGY  FH+KLA LGA I R
Sbjct: 357 GPVQLQGAEVAATDLRAGAALILAGLVAEGHTRVTELKHLDRGYVDFHQKLAALGADIER 416

Query: 420 SSE 422
            ++
Sbjct: 417 VND 419
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 363/422 (86%), Positives = 391/422 (92%)

Query: 5   MDKIIVEGGQTQLQGQVVIEGAKNAVLPLLAATILPSQGKTLLTNVPILSDVFTMNNVVR 64
           +DKII+EGGQT+L+G+VVIEGAKNAVLPLLAA+ILPS+GKT+L NVPILSDVFTMNNVVR
Sbjct: 1   VDKIIIEGGQTRLEGEVVIEGAKNAVLPLLAASILPSKGKTILRNVPILSDVFTMNNVVR 60

Query: 65  GLDIQVDFNCDKKEILVDASGDILDVAPYEFVSQMRASIVVLGPILARNGHAKVSMPGGC 124
           GLDI+VDFN    EI VDASG ILD APYE+VSQMRASIVVLGPILARNGHAKVSMPGGC
Sbjct: 61  GLDIRVDFNEAANEITVDASGHILDEAPYEYVSQMRASIVVLGPILARNGHAKVSMPGGC 120

Query: 125 TIGSRPIDLHLKGLEAMGATITQNGGDITAQAEKLKGANIYMDFPSVGATQNLMMAATLA 184
           TIGSRPI+LHLKGLEAMGATITQ GGDITAQA++L+GA IYMDFPSVGATQNLMMAATLA
Sbjct: 121 TIGSRPINLHLKGLEAMGATITQKGGDITAQADRLQGAMIYMDFPSVGATQNLMMAATLA 180

Query: 185 SGTTTIENAAREPEIVDLAQLLNKMGAKVKGAGTETLTIIGVDALHGTEHDVVQDRIEAG 244
            G TTIENAAREPEIVDLAQ LNKMGA+++GAGTETLTI GV  L G EHDVVQDRIEAG
Sbjct: 181 DGVTTIENAAREPEIVDLAQFLNKMGARIRGAGTETLTITGVTHLRGVEHDVVQDRIEAG 240

Query: 245 TFMVAAAMTSGNVLVKDAIWEHNRPLISKLMEMGVEVSEEEDGIRVKADTKKLKPVTVKT 304
           TFMVAAAMTSGNVL++DA+WEHNRPLISKLMEMGV V+EEE GIRV+A+T KLKPVTVKT
Sbjct: 241 TFMVAAAMTSGNVLIRDAVWEHNRPLISKLMEMGVSVTEEEYGIRVQANTPKLKPVTVKT 300

Query: 305 LPHPGFPTDMQAQFTALMAVVNGESTMIETVFENRFQHLEEMRRMGLQTEILRDTAMIHG 364
           LPHPGFPTDMQAQFTALMAVVNGESTM+ETVFENRFQHLEEMRRMGLQ+EILR+TAMIHG
Sbjct: 301 LPHPGFPTDMQAQFTALMAVVNGESTMVETVFENRFQHLEEMRRMGLQSEILRETAMIHG 360

Query: 365 GRALQGAPVMSTDLRASAALILAGMVAQGQTVVGQLTHLDRGYYQFHEKLAALGANIKRVSE 426
           GR LQGAPVMSTDLRASAALIL G+VAQG T+V  L HLDRGYYQFHEKLA LGA I RSSE
Sbjct: 361 GRQLQGAPVMSTDLRASAALILTGIVAQGVTIVNNLVHLDRGYYQFHEKLAKLGATISRSSE 422
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1409

A DNA sequence (GBSx1494) was identified in *S.agalactiae* <SEQ ID 4323> which encodes the amino acid sequence <SEQ ID 4324>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2096(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23756 GB:AB009314 proton-translocating ATPase, epsiron
           subunit [Streptococcus bovis]
 Identities = 102/138 (73%), Positives = 121/138 (86%), Gaps = 1/138 (0%)

Query: 1   MAQLTVQVVTPDGIRYDHHASLITVRTPDGEMGILPGHINLIAPLNVHQMKINRSHQEG- 59
           M  +TVQVVTPDGIRYDHHA+ I+V+TPDGEMGILP HINLIAPL VH+MKI+R+
Sbjct: 1   MTFMTVQVVTPDGIRYDHHANFISVKTPDGEMGILPEHINLIAPLTVHEMKIHRTDDPNH 60

Query: 60  VDWVAVNGGIIEVNEDQVTIVADSAERARDIDLNRAERAKERAERALEKAQTTQNIDEMR 119
           VDWVA+NGGIIE+ ++ VTIVADSAER RDID++RAERAK RAER LE+AQ+T +IDE+R
Sbjct: 61  VDWVAINGGIIEIKDNLVTIVADSAERERDIDVSRAERAKIRAERKLEQAQSTHDIDEVR 120

Query: 120 RAEVALRRAINRISVGKK 137
           RA+VALRRA+NRISVG K
Sbjct: 121 RAQVALRRALNRISVGNK 138
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4325> which encodes the amino acid sequence <SEQ ID 4326>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2539(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 100/138 (72%), Positives = 119/138 (85%), Gaps = 1/138 (0%)

Query: 1   MAQLTVQVVTPDGIRYDHHASLITVRTPDGEMGILPGHINLIAPLNVHQMKINRSHQ-EG 59
           M Q+TVQVVTPDGI+YDHHA  I+V TPDGEMGILP HINLIAPL VH+MKI R  + E
Sbjct: 1   MTQMTVQVVTPDGIKYDHHAKFISVTTPDGEMGILPNHINLIAPLQVHEMKIRRGGEDEK 60

Query: 60  VDWVAVNGGIIEVNEDQVTIVADSAERARDIDLNRAERAKERAERALEKAQTTQNIDEMR 119
           VDW+A+NGGIIE+ ++ VTIVADSAER RDID++RAERAK RAER + +A+TT NIDE+R
Sbjct: 61  VDWIAINGGIIEIKDNVVTIVADSAERDRDIDVSRAERAKLRAEREIAQAETTHNIDEVR 120

Query: 120 RAEVALRRAINRISVGKK 137
           RA+VALRRA+NRI+V KK
Sbjct: 121 RAKVALRRALNRINVSKK 138
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1410

A DNA sequence (GBSx1495) was identified in *S.agalactiae* <SEQ ID 4327> which encodes the amino acid sequence <SEQ ID 4328>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein is similar to the beta subunit of the *S.mutans* ATPase:

```
>GP:AAD13383 GB:U31170 ATPase, beta subunit [Streptococcus mutans]
 Identities = 435/466 (93%), Positives = 455/466 (97%)

Query: 1   MSSGKIAQVVGPVVDVVFASGDKLPEINNALIVYKNGDKSQKVVLEVALELGDGLVRTIA 60
            MS+GKIAQVVGPVVDV FA+ DKLPEINNAL+VYK+GDKSQ++VLEVALELGDGLVRTIA
Sbjct: 1   MSTGKIAQVVGPVVDVAFATDDKLPEINNALVVYKDGDKSQRIVLEVALELGDGLVRTIA 60

Query: 61  MESTDGLTRGLEVLDTGRAISVPVGKDTLGRVFNVLGDAIDLEEPFAEDAERQPIHKKAP 120
            MESTDGLTRGLEV DTGRAISVPVGK+TLGRVFNVLGD IDL++PFAEDAERQPIHKKAP
Sbjct: 61  MESTDGLTRGLEVFDTGRAISVPVGKETLGRVFNVLGDTIDLDKPFAEDAERQPIHKKAP 120

Query: 121 SFDELSTSSEILETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIAQEHGGIS 180
            SFD+LSTS+EILETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIAQEHGGIS
Sbjct: 121 SFDDLSTSTEILETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIAQEHGGIS 180

Query: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240
            VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE
Sbjct: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240

Query: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATEMGQLQERITSTKKGSVTSI 300
            GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATEMGQLQERITSTKKGSVTSI
Sbjct: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATEMGQLQERITSTKKGSVTSI 300

Query: 301 QAIYVPADDYTDPAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRALTPEIVGDEH 360
            QAIYVPADDYTDPAPATAFAHLDSTTNLER+LTQMGIYPAVDPLASSSRAL+PEIVG EH
Sbjct: 301 QAIYVPADDYTDPAPATAFAHLDSTTNLERRLTQMGIYPAVDPLASSSRALSPEIVGQEH 360

Query: 361 YEVATEVQRVLQRYRELQDIIAILGMDELSDEEKTLVGRARRIQFFLSQNFNVAETFTGQ 420
            Y+VATEVQ VLQRYRELQDIIAILGMDELSDEEKTLVGRARRIQFFLSQNFNVAE FTGQ
Sbjct: 361 YDVATEVQHVLQRYRELQDIIAILGMDELSDEEKTLVGRARRIQFFLSQNFNVAEQFTGQ 420

Query: 421 PGSYVPVEETVRGFKEILDGKHDQIPEDAFRMVGGIEDVIAKAEKM 466
            PGSYVPV ETVRGFKEIL+GK+D++PEDAFR VG IEDV+ KA+KM
Sbjct: 421 PGSYVPVAETVRGFKEILEGKYDELPEDAFRSVGAIEDVVEKAKKM 466
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4329> which encodes the amino acid sequence <SEQ ID 4330>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0275(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 440/468 (94%), Positives = 456/468 (97%)

Query: 1   MSSGKIAQVVGPVVDVVFASGDKLPEINNALIVYKNGDKSQKVVLEVALELGDGLVRTIA 60
```

```
            MSSGKIAQVVGPVVDV+FASGDKLPEINNALIVYK+ DK QK+VLEVALELGDG+VRTIA
Sbjct: 1    MSSGKIAQVVGPVVDVMFASGDKLPEINNALIVYKDSDKKQKIVLEVALELGDGMVRTIA 60

Query: 61   MESTDGLTRGLEVLDTGRAISVPVGKDTLGRVFNVLGDAIDLEEPFAEDAERQPIHKKAP 120
            MESTDGLTRGLEVLDTGRAISVPVGK+TLGRVFNVLG+ IDLEEPFAED +RQPIHKKAP
Sbjct: 61   MESTDGLTRGLEVLDTGRAISVPVGKETLGRVFNVLGETIDLEEPFAEDVDRQPIHKKAP 120

Query: 121  SFDELSTSSEILETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIAQEHGGIS 180
            SFDELSTSSEILETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIAQEHGGIS
Sbjct: 121  SFDELSTSSEILETGIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIAQEHGGIS 180

Query: 181  VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240
            VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE
Sbjct: 181  VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240

Query: 241  GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATEMGQLQERITSTKKGSVTSI 300
            GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATEMGQLQERITST+KGSVTSI
Sbjct: 241  GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATEMGQLQERITSTQKGSVTSI 300

Query: 301  QAIYVPADDYTDPAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRALTPEIVGDEH 360
            QAIYVPADDYTDPAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRAL+PEIVG+EH
Sbjct: 301  QAIYVPADDYTDPAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRALSPEIVGEEH 360

Query: 361  YEVATEVQRVLQRYRELQDIIAILGMDELSDEEKTLVGRARRIQFFLSQNFNVAETFTGQ 420
            Y VATEVQRVLQRYRELQDIIAILGMDELSDEEKTLVGRARRIQFFLSQNFNVAE FTG
Sbjct: 361  YAVATEVQRVLQRYRELQDIIAILGMDELSDEEKTLVGRARRIQFFLSQNFNVAEQFTGL 420

Query: 421  PGSYVPVEETVRGFKEILDGKHDQIPEDAFRMVGGIEDVIAKAEKMNY 468
            PGSYVPV +TVRGFKEIL+GK+D++PEDAFR VG IEDVI KAEKM +
Sbjct: 421  PGSYVPVADTVRGFKEILEGKYDELPEDAFRSVGPIEDVIKKAEKMGF 468
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1411

A DNA sequence (GBSx1496) was identified in *S.agalactiae* <SEQ ID 4331> which encodes the amino acid sequence <SEQ ID 4332>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1889(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23754 GB:AB009314 proton-translocating ATPase, gamma subunit
           [Streptococcus bovis]
 Identities = 252/293 (86%), Positives = 278/293 (94%), Gaps = 2/293 (0%)

Query: 1    MAGSLSEIKDKILSTEKTSKITSAMQMVSSAKLVKSEQAARDFQVYASKIRQITTNLLKS 60
            MAGSLSEIK KI+ST+KTS IT AMQMVS+AKL KSEQAA+DFQVYASKIRQITT+LLKS
Sbjct: 1    MAGSLSEIKGKIISTQKTSHITGAMQMVSAAKLTKSEQAAKDFQVYASKIRQITTDLLKS 60

Query: 61   DLVSGSDNPMLSSRPVKKTGYIVITSDKGLVGGYNSKILKAMMDTITDYHTENDDYAIIS 120
            +LV+GS NPML++RPVKKTGYIVITSDKGLVGGYNSKILKAMMD I +YH ++ +YAII+
Sbjct: 61   ELVNGSKNPMLAARPVKKTGYIVITSDKGLVGGYNSKILKAMMDLIEEYH-QDGNYAIIA 119

Query: 121  IGSVGSDFFKARGMNVSFELRGLEDQPSFDQVGKIIAQAVEMYKNELFDELYVCYNHHVN 180
            IG +G+DFFKARGMNV FELRGLEDQPSF+QVG IIA++VEMYKNELFDELYVCYNHHVN
Sbjct: 120  IGGIGADFFKARGMNVVFELRGLEDQPSFEQVGNIIAKSVEMYKNELFDELYVCYNHHVN 179

Query: 181  SLTSQVRMQQMLPIKELDAEEASEDRVITGFELEPNREVILEQLLPQYTESLIYGAIIDA 240
```

```
           SLTSQVR+QQMLPI ELDA+EA+E+ V +GFELEPNRE+ILEQLLPQYTESLIYGAI+DA
Sbjct: 180 SLTSQVRVQQMLPIAELDADEAAEEGV-SGFELEPNREMILEQLLPQYTESLIYGAIVDA 238

Query: 241 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRARQAAITQEITEIVAGANALE 293
           KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRARQAAITQEITEIVAGANALE
Sbjct: 239 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRARQAAITQEITEIVAGANALE 291
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4333> which encodes the amino acid sequence <SEQ ID 4334>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1969(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 251/293 (85%), Positives = 275/293 (93%), Gaps = 2/293 (0%)

Query: 1   MAGSLSEIKDKILSTEKTSKITSAMQMVSSAKLVKSEQAARDFQVYASKIRQITTNLLKS 60
           MAGSLSEIK KI+STEKTSKITSAM+MVSSAKLVKSEQAARDFQ+YASKIRQITT+LLKS
Sbjct: 1   MAGSLSEIKAKIISTEKTSKITSAMRMVSSAKLVKSEQAARDFQIYASKIRQITTDLLKS 60

Query: 61  DLVSGSDNPMLSSRPVKKTGYIVITSDKGLVGGYNSKILKAMMDTITDYHTENDDYAIIS 120
           +L  GSDNPML SRPVKKTGYIVITSDKGLVGGYNSKILK++MD IT+YH +  DY IIS
Sbjct: 61  ELTIGSDNPMLVSRPVKKTGYIVITSDKGLVGGYNSKILKSVMDMITEYHADG-DYEIIS 119

Query: 121 IGSVGSDFFKARGMNVSFELRGLEDQPSFDQVGKIIAQAVEMYKNELFDELYVCYNHHVN 180
           IGSVGSDFFKARGMNV+FELRGL DQPSF+QV +II+Q+V+M+ NE+FDELYVCYNHHVN
Sbjct: 120 IGSVGSDFFKARGMNVAFELRGLADQPSFEQVRQIISQSVDMFVNEIFDELYVCYNHHVN 179

Query: 181 SLTSQVRMQQMLPIKELDAEEASEDRVITGFELEPNREVILEQLLPQYTESLIYGAIIDA 240
           SLTSQVR+QQMLPI +L A+EA+E+ V TGFELEPNR  IL+QLLPQ+TESLIYGAIIDA
Sbjct: 180 SLTSQVRVQQMLPISDLVADEAAEEGV-TGFELEPNRHDILDQLLPQFTESLIYGAIIDA 238

Query: 241 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRARQAAITQEITEIVAGANALE 293
           KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRARQAAITQEITEIVAGANALE
Sbjct: 239 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRARQAAITQEITEIVAGANALE 291
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1412

A DNA sequence (GBSx1497) was identified in *S.agalactiae* <SEQ ID 4335> which encodes the amino acid sequence <SEQ ID 4336>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1963(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1413

A DNA sequence (GBSx1498) was identified in *S.agalactiae* <SEQ ID 4337> which encodes the amino acid sequence <SEQ ID 4338>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3146(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein is similar to the alpha subunit of the proton-translocating ATPase from *S.bovis*:

```
>GP:BAA23753 GB:AB009314 proton-translocating ATPase, alpha subunit
          [Streptococcus bovis]           Length = 501
 Identities = 482/501 (96%), Positives = 497/501 (98%)

Query: 1   MAINAQEISALIKKQIEDFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60
            MAINAQEISALIKKQIE+FQPNFDVTETG+VTYIGDGIARARGLDNAMSGELLEFSNGA+
Sbjct: 1   MAINAQEISALIKKQIENFQPNFDVTETGVVTYIGDGIARARGLDNAMSGELLEFSNGAF 60

Query: 61  GMAQNLESNDVGIIILGDFSEIREGDVVKRTGKIMEVPVGEAMIGRVVNPLGQPVDGLGE 120
            GMAQNLESNDVGIIILGDFS IREGD VKRTGKIMEVPVGEA+IGRVVNPLGQPVDGLG+
Sbjct: 61  GMAQNLESNDVGIIILGDFSTIREGDEVKRTGKIMEVPVGEALIGRVVNPLGQPVDGLGD 120

Query: 121 IETTATRPVETPAPGVMQRKSVFEPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI 180
            I+TTATRPVETPAPGVMQRKSV EPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI
Sbjct: 121 IKTTATRPVETPAPGVMQRKSVSEPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI 180

Query: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLFIAPY 240
            DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLL+IAPY
Sbjct: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLYIAPY 240

Query: 241 AGVAMAEEFMYNGKHVLIVYDDLSKQAVAYRELSLLLRRPPGREAYPGDVFYLHSRLLER 300
            AGVAMAEEFMYNGKHVLIVYDDLSKQAVAYRELSLLLRRPPGREAYPGDVFYLHSRLLER
Sbjct: 241 AGVAMAEEFMYNGKHVLIVYDDLSKQAVAYRELSLLLRRPPGREAYPGDVFYLHSRLLER 300

Query: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360
            SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG
Sbjct: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360

Query: 361 SSVSRVGGAAQIKAMKRVAGTLRLDLASYRELEAFTQFGSDLDAATQAKLNRGRRTVEVL 420
```

```
 bacterial cytoplasm --- Certainty=0.3654(Affirmative) < succ>
  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 477/501 (95%), Positives = 490/501 (97%)

Query: 1   MAINAQEISALIKKQIEDFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60
           +AINAQEISALIKKQIE+FQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEF NGAY
Sbjct: 1   LAINAQEISALIKKQIENFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFENGAY 60

Query: 61  GMAQNLESNDVGIIILGDFSEIREGDVVKRTGKIMEVPVGEAMIGRVVNPLGQPVDGLGE 120
           GMAQNLESNDVGIIILGDFS IREGDVVKRTGKIMEVPVGEA+IGRVVNPLGQPVDGLG+
Sbjct: 61  GMAQNLESNDVGIIILGDFSAIREGDVVKRTGKIMEVPVGEALIGRVVNPLGQPVDGLGD 120

Query: 121 IETTATRPVETPAPGVMQRKSVFEPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI 180
           IETT  RPVETPAPGVMQRKSV EPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI
Sbjct: 121 IETTGFRPVETPAPGVMQRKSVSEPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI 180

Query: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLFIAPY 240
           DAILNQKGQDMICIYVAIGQKESTVRTQVETLR+YGALDYTIVVTASASQPSPLLFIAPY
Sbjct: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRRYGALDYTIVVTASASQPSPLLFIAPY 240

Query: 241 AGVAMAEEFMYNGKHVLIVYDDLSKQAVAYRELSLLLRRPPGREAYPGDVFYLHSRLLER 300
           AGVAMAEEFMY GKHVLIVYDDLSKQAVAYRELSLLLRRPPGREAYPGDVFYLHSRLLER
Sbjct: 241 AGVAMAEEFMYQGKHVLIVYDDLSKQAVAYRELSLLLRRPPGREAYPGDVFYLHSRLLER 300

Query: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360
           SAKVSD LGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG
Sbjct: 301 SAKVSDDLGGGSITALPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360

Query: 361 SSVSRVGGAAQIKAMKRVAGTLRLDLASYRELEAFTQFGSDLDAATQAKLNRGRRTVEVL 420
           SSVSRVGG+AQIKAMK+VAGTLRLDLASYRELEAFTQFGSDLDAATQAKLNRGRRTVE+L
Sbjct: 361 SSVSRVGGSAQIKAMKKVAGTLRLDLASYRELEAFTQFGSDLDAATQAKLNRGRRTVEIL 420

Query: 421 KQPLHKPLPVEKQVVILYALTHGFLDDVPVNDILAFEEALYDYFDAHYDNLFETIRTTKD 480
           KQPLHKPLPVEKQVVILYALTHGFLDDVPV+DILAFEEALYDYFD HY++LFETIRTTKD
Sbjct: 421 KQPLHKPLPVEKQVVILYALTHGFLDDVPVDDILAFEEALYDYFDVHYNDLFETIRTTKD 480

Query: 481 LPEEAELDAAIQAFKDQSQFK 501
           LPEEA LDAAI+AFK+ S FK
Sbjct: 481 LPEEAALDAAIKAFKEHSNFK 501
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1414

A DNA sequence (GBSx1499) was identified in *S.agalactiae* <SEQ ID 4341> which encodes the amino acid sequence <SEQ ID 4342>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1896(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23752 GB:AB009314 proton-translocating ATPase, delta subunit
            [Streptococcus bovis]
 Identities = 98/178 (55%), Positives = 127/178 (71%)
```

```
Query: 1   MNKKTQALIEQYSKSLVEVAIEHKIVEKIQQEVAALIDIFETSELEGVLSSLAVSHDEKQ 60
           M+KKTQAL+EQY+KSLVE+AIE   + ++Q E  AL+ +FE + L   LSSL VS DEK
Sbjct: 1   MDKKTQALVEQYAKSLVEIAIEKDSLAELQSETEALLSVFEETNLADFLSSLVVSRDEKV 60

Query: 61  HFVKTLQTSCSTYLVNFLEVIVQNEREALLYPILKSVDQELIKVNGQYPIQITTAVALSP 120
             V+ LQ S S Y+ NFLEVI+QNEREA L  IL+ V ++ +    Q+ I +TTAVAL+
Sbjct: 61  KLVRLLQESSSVYMNNFLEVILQNEREAFLKAILEGVQKDFVIATNQHDIVVTTAVALTD 120

Query: 121 EQKERLFDIAKTKLALPNGQLVEHIDPSIVGGFVVNANNKVIDASVRNQLHQFKMKLK 178
           EQKER+  +   K  +  G+LVE+ID SI+GGFV+N NNKVID S+R QL +FKM LK
Sbjct: 121 EQKERILALVAEKFGVKAGKLVENIDESILGGFVINVNNKVIDTSIRRQLQEFKMNLK 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4343> which encodes the amino acid sequence <SEQ ID 4344>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1668(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 86/178 (48%), Positives = 125/178 (69%)

Query: 1   MNKKTQALIEQYSKSLVEVAIEHKIVEKIQQEVAALIDIFETSELEGVLSSLAVSHDEKQ 60
           M KK QALIEQY+KSLVEVA EH  ++ +Q +V A+++ F T+ L+  LSS AV H EK
Sbjct: 1   MTKKEQALIEQYAKSLVEVASEHHSLDALQADVLAILETFVTTNLDQSLSSQAVPHAEKI 60

Query: 61  HFVKTLQTSCSTYLVNFLEVIVQNEREALLYPILKSVDQELIKVNGQYPIQITTAVALSP 120
             +  L+ + S Y+ NFL +I+QNEREA LY +L++V  E+  V+ QY + +T+++ L+
Sbjct: 61  KLLTLLKGNNSVYMNNFLNLILQNEREAYLYQMLQAVLNEIAIVSNQYDVTVTSSLPLTE 120

Query: 121 EQKERLFDIAKTKLALPNGQLVEHIDPSIVGGFVVNANNKVIDASVRNQLHQFKMKLK 178
           EQK R+  +   K A+  G+L+E +DPS++GGF+++ NNKVID S+R QL  FKM LK
Sbjct: 121 EQKSRVRAVVAKKFAVTAGRLIEKVDPSLIGGFIISVNNKVIDTSIRRQLQAFKMNLK 178
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1415

A DNA sequence (GBSx1500) was identified in *S.agalactiae* <SEQ ID 4345> which encodes the amino acid sequence <SEQ ID 4346>. This protein is predicted to be ATP synthase b chain (atpF). Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]
 Identities = 103/165 (62%), Positives = 130/165 (78%)

Query: 1   MSILINSTTIGDIIIVSGSVLLLFILIKTFAWKQITGIFEAREQKIANDIDTAEQARQQA 60
           MS LIN T++G+++IV+GS +LL +L+K FAW Q+  IF+ RE+KIA DID AE +RQ A
Sbjct: 1   MSTLINGTSLGNLLIVTGSFILLLLLVKKFAWSQLAAIFKTREEKIAKDIDDAENSRQNA 60
```

```
Query: 61  EAFATKREEELSNAKTEANQIIDNAKETGLAKGDQIISEAKTEADRLKEKAHQDIAQNKA 120
           +    KR+ EL+ AK EA QIIDNAKETG A+  +II+EA  EA RLK+KA+QDIA +KA
Sbjct: 61  QVLENKRQVELNQAKDEAAQIIDNAKETGKAQESKIITEAHEEAGRLKDKANQDIATSKA 120

Query: 121 EALADVKGEVADLTVLLAEKIMVSNLDKEAQSNLIDSYIKKLGDA 165
           EAL+ VK +VADL+VLLAEKIM  NLDK AQ +LIDSY+ KLGDA
Sbjct: 121 EALSSVKADVADLSVLLAEKIMAKNLDKTAQGDLIDSYLDKLGDA 165
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4347> which encodes the amino acid sequence <SEQ ID 4348>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]
 Identities = 88/159 (55%), Positives = 122/159 (76%)

Query: 6   GELVGNFILVTGSVIVLLLLIKKFAWGAIESILQTRSQQISRDIDQAEQSRLSAQQLEAK 65
           G  +GN ++VTGS I+LLLL+KKFAW  + +I +TR ++I++DID AE SR +AQ LE K
Sbjct: 7   GTSLGNLLIVTGSFILLLLLVKKFAWSQLAAIFKTREEKIAKDIDDAENSRQNAQVLENK 66

Query: 66  SQANLDASRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAV 125
            Q  L+ ++ +A++II +AKE G+ Q  K++ EA +EA RLK+KA  DI  SK++A+S+V
Sbjct: 67  RQVELNQAKDEAAQIIDNAKETGKAQESKIITEAHEEAGRLKDKANQDIATSKAEALSSV 126

Query: 126 KTEMSDLTVLLAEKIMGANLDKTAQSQLIDSYLDDLGEA 164
           K +++DL+VLLAEKIM  NLDKTAQ  LIDSYLD LG+A
Sbjct: 127 KADVADLSVLLAEKIMAKNLDKTAQGDLIDSYLDKLGDA 165
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 81/156 (51%), Positives = 115/156 (72%)

Query: 10  IGDIIIVSGSVLLLFILIKTFAWKQITGIFEAREQKIANDIDTAEQARQQAEAFATKREE 69
           +G+ I+V+GSV++L +LIK FAW  I  I + R Q+I+ DID AEQ+R  A+    K +
Sbjct: 9   VGNFILVTGSVIVLLLLIKKFAWGAIESILQTRSQQISRDIDQAEQSRLSAQQLEAKSQA 68

Query: 70  ELSNAKTEANQIIDNAKETGLAKGDQIISEAKTEADRLKEKAHQDIAQNKAEALADVKGE 129
            L  ++ +A++II +AKE G  +GD++++EA  EA RLKEKA  DI Q+K++A++ VK E
Sbjct: 69  NLDASRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAVKTE 128

Query: 130 VADLTVLLAEKIMVSNLDKEAQSNLIDSYIKKLGDA 165
           ++DLTVLLAEKIM +NLDK AQS LIDSY+  LG+A
Sbjct: 129 MSDLTVLLAEKIMGANLDKTAQSQLIDSYLDDLGEA 164
```

SEQ ID 4346 (GBS169) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 6; MW 18kDa).

The GBS169-His fusion product was purified (Figure 200, lane 11) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 250). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1416

A DNA sequence (GBSx1501) was identified in *S.agalactiae* <SEQ ID 4349> which encodes the amino acid sequence <SEQ ID 4350>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.73   Transmembrane    20 -  36 (  14 -  42)
   INTEGRAL    Likelihood = -5.20   Transmembrane   207 - 223 ( 206 - 228)
   INTEGRAL    Likelihood = -4.35   Transmembrane    78 -  94 (  73 -  97)
   INTEGRAL    Likelihood = -4.09   Transmembrane   113 - 129 ( 113 - 133)
   INTEGRAL    Likelihood = -2.39   Transmembrane   174 - 190 ( 174 - 190)

----- Final Results -----
               bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23750 GB:AB009314 proton-translocating ATPase, a subunit
           [Streptococcus bovis]
 Identities = 149/238 (62%), Positives = 180/238 (75%)

Query: 1   MESTSNPTVSFLGIDFDLTILAMSLLTITIIFILVFWASRKMTIKPKGKQNVLEYVYELV 60
           ME++ NPT    GI+FDLTILAMSLLT+ I F ++FWA+RKMT+KPKGKQN +EYVYE V
Sbjct: 1   METSVNPTAHVFGIEFDLTILAMSLLTVIISFGIIFWATRKMTLKPKGKQNFIEYVYEFV 60

Query: 61  NNTISQNLGHYTKNYSLLMFILFSFVFIANNLGLMTSLKTHEHNFWTSPTANFGVDITLS 120
            NTI  NLG YT  YSLLMF  F F+ IANNLGL+  L++ ++NFWTSPT+   VD T S
Sbjct: 61  QNTIKPNLGEYTPKYSLLMFTFFFFILIANNLGLLVKLESEDYNFWTSPTSTIMVDCTWS 120

Query: 121 LLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPMNLLEEVTNVASLALRLFGNIFSGEVV 180
           L+VA + H+EG+RKKG+  YLKG+LSP P MLPMN+LE+ TNV SLALRLFGNI++GEVV
Sbjct: 121 LIVAIVVHVEGVRKKGVKAYLKGYLSPFPMMLPMNILEQFTNVLSLALRLFGNIYAGEVV 180

Query: 181 TGLLLQLAVLSPFTGPLAFALNIVWTAFSMFIGFIQAYVFIILSSSYIGHKVHGDEEE 238
           T L++     S    P A ALN+ W AFS FIG IQAYVF ILSS YI  K+  DE+E
Sbjct: 181 TALIVGFGTKSLIFAPFALALNLAWVAFSAFIGCIQAYVFTILSSKYISEKLPEDEDE 238
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4351> which encodes the amino acid sequence <SEQ ID 4352>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.73   Transmembrane    79 -  95 (  72 -  97)
   INTEGRAL    Likelihood = -4.35   Transmembrane   115 - 131 ( 112 - 132)
   INTEGRAL    Likelihood = -2.13   Transmembrane   200 - 216 ( 197 - 216)

----- Final Results -----
               bacterial membrane --- Certainty=0.2890(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 124/239 (51%), Positives = 169/239 (69%), Gaps = 3/239 (1%)

Query: 1   MESTSNPTVSFLGIDFDLTILAMSLLTITIIFILVFWASRKMTIKPKGKQNVLEYVYELV 60
           ME    P +    I F+LT+LA+ ++TI I+F  VFWASR+M +KP+GKQ  LEY+   V
Sbjct: 1   MEEAKIPMLKLGPITFNLTLLAVCIVTIAIVFAFVFWASRQMKLKPEGKQTALEYLISFV 60

Query: 61  NNTISQNLGH-YTKNYSLLMFILFSFVFIANNLGLMTSLKT-HEHNFWTSPTANFGVDIT 118
           +    ++L H   K+YSLL+F +F FV +ANNLGL T L+T + +N WTSPTAN   D+
Sbjct: 61  DGIGEEHLDHNLQKSYSLLLFTIFLFVAVANNLGLFTKLETVNGYNLWTSPTANLAFDLA 120
```

```
Query: 119 LSLLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPMNLLEEVTNVASLALRLFGNIFSGE 178
           LSL +  + HIEG+R++G+  +LK   +P P M PMNLLEE TN  SLA+RLFGNIF+GE
Sbjct: 121 LSLFITLMVHIEGVRRRGLVAHLKRLATPWP-MTPMNLLEEFTNFLSLAIRLFGNIFAGE 179

Query: 179 VVTGLLLQLAVLSPFTGPLAFALNIVWTAFSMFIGFIQAYVFIILSSSYIGHKVHGDEE 237
           VVTGL++QLA    +  P+AF +N+ WTAFS+FI  IQA+VF  L+++Y+G KV+  EE
Sbjct: 180 VVTGLIVQLANYRVYWWPIAFLVNMAWTAFSVFISCIQAFVFTKLTATYLGKKVNESEE 238
```

A related GBS gene <SEQ ID 8803> and protein <SEQ ID 8804> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 1
McG: Discrim Score:     -3.50
GvH: Signal Score (-7.5): -3.36
     Possible site: 29
>>> Seems to have no N-terminal signal sequence
ALOM program  count: 5 value: -11.73 threshold:  0.0
   INTEGRAL    Likelihood =-11.73  Transmembrane   20 -  36 (  14 -  42)
   INTEGRAL    Likelihood = -5.20  Transmembrane  207 - 223 ( 206 - 228)
   INTEGRAL    Likelihood = -4.35  Transmembrane   78 -  94 (  73 -  97)
   INTEGRAL    Likelihood = -4.09  Transmembrane  113 - 129 ( 113 - 133)
   INTEGRAL    Likelihood = -2.39  Transmembrane  174 - 190 ( 174 - 190)
   PERIPHERAL  Likelihood =  5.30     156
 modified ALOM score:   2.85

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01818(301 - 1014 of 1314)
GP|2662321|dbj|BAA23750.1||AB009314(1 - 238 of 239) proton-translocating ATPase, a subunit
{Streptococcus bovis}
%Match = 35.0
%Identity = 62.2  %Similarity = 78.6
Matches = 148  Mismatches = 51  Conservative Sub.s = 39

204       234       264       294       324       354       384       414
XANCQTLMLPGVGFIERYFLRSICVYILSKIDDNLEKKEG*GLESTSNPTVSFLGIDFDLTILAMSLLTITIIFILVFWA
                                          :|:: |||  :||:||||||||||||: | | ::|||
                                          METSVNPTAHVFGIEFDLTILAMSLLTVIISFGIIFWA
                                                  10        20        30

444       474       504       534       564       594       624       654
SRKMTIKPKGKQNVLEYVYELVNNTISQNLGHYTKNYSLLMFILFSFVFIANNLGLMTSLKTHEHNFWTSPTANFGVDIT
:||||:||||||| :|||||:| |||  ||| ||  |||||| :| |::|||||||:  |:: ::|||||||:   || |
TRKMTLKPKGKQNFIEYVYEFVQNTIKPNLGEYTPKYSLLMFTFFFFILIANNLGLLVKLESEDYNFWTSPTSTIMVDCT
          50        60        70        80        90       100       110

684       714       744       774       804       834       864       894
LSLLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPMNLLEEVTNVASLALRLFGNIFSGEVVTGLLLQLAVLSPFTGPLA
 ||:|| : |:||:||||:  ||||:||| | |||||:||: ||| ||||||||||::||||| |:: :   |    |:|
WSLIVAIVVHVEGVRKKGVKAYLKGYLSPFPMMLPMNILEQFTNVLSLALRLFGNIYAGEVVTALIVGFGTKSLIFAPFA
         130       140       150       160       170       180       190

924       954       984       1014      1044      1074      1104      1134
FALNIVWTAFSMFIGFIQAYVFIILSSSYIGHKVHGDEEE*EKRGEICQYLLIVQRLVISLSYLALCFSYLS*LRLLHGN
:|||: | ||| ||| |||||| |||| ||  |:  ||:|
LALNLAWVAFSAFIGCIQAYVFTILSSKYISEKLPEDEDET
         210       220       230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1417

A DNA sequence (GBSx1502) was identified in *S.agalactiae* <SEQ ID 4353> which encodes the amino acid sequence <SEQ ID 4354>. This protein is predicted to be ATP synthase c subunit (atpE). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.62    Transmembrane   48 -  64 (  42 -  65)

----- Final Results -----
               bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23749 GB:AB009314 proton-translocating ATPase, c subunit [Streptococcus bovis]
 Identities = 56/65 (86%), Positives = 59/65 (90%)

Query: 1  MNLAILALGFAVMGVSIGEGILVANIAKSAARQPEMFSKLQTLMFTGVAFIEGTFFVLFA 60
           +NL ILALG AV+GVS+GEGILVANIAKSAARQPEMFSKLQTLMF GVAFIEGTFFVL A
Sbjct: 2  LNLKILALGLAVLGVSLGEGILVANIAKSAARQPEMFSKLQTLMFLGVAFIEGTFFVLLA 61

Query: 61 FTFLV 65
            TF V
Sbjct: 62 STFFV 66
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4355> which encodes the amino acid sequence <SEQ ID 4356>. Analysis of this protein sequence reveals the following:

```
     Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.26    Transmembrane   47 -  63 (  41 -  64)

----- Final Results -----
               bacterial membrane --- Certainty=0.3102(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00920 GB:AF001955 UncE [Streptococcus sanguinis]
 Identities = 50/66 (75%), Positives = 58/66 (87%), Gaps = 1/66 (1%)

Query: 1  MNPIF-ALALACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLMILGVAFIEGTFFVTLV 59
           MN  F  L  ACFGVS+AEG +M+NLFKAASRQPEIIGQLRSL+ILG+AF+EGTFFVTL
Sbjct: 1  MNLTFLGLCFACFGVSIAEGLIMSNLFKAASRQPEIIGQLRSLLILGIAFVEGTFFVTLA 60

Query: 60 MAFILK 65
           MAF++K
Sbjct: 61 MAFVIK 66
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 33/62 (53%), Positives = 45/62 (72%)

Query: 5  ILALGFAVMGVSIGEGILVANIAKSAARQPEMFSKLQTLMFTGVAFIEGTFFVLFAFTFLVR 66
           I AL  A  GVS+ EG L+AN+ K+A+RQPE+  +L++LM  GVAFIEGTFFV     F+++
Sbjct: 4  IFALALACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLMILGVAFIEGTFFVTLVMAFILK 65
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1418

A DNA sequence (GBSx1503) was identified in *S.agalactiae* <SEQ ID 4357> which encodes the amino acid sequence <SEQ ID 4358>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2562(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1419

A DNA sequence (GBSx1504) was identified in *S.agalactiae* <SEQ ID 4359> which encodes the amino acid sequence <SEQ ID 4360>. This protein is predicted to be bacterial glycogen synthase (glgA). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1574(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA19591 GB:D87026 bacterial glycogen synthase [Bacillus
           stearothermophilus]
 Identities = 220/475 (46%), Positives = 312/475 (65%), Gaps = 1/475 (0%)

Query: 1   MKIMFVAAEGAPFAKTGGLGDVIGALPKSLSKKGHDVAVVMPYYDMVDQKFGDQIENLMY 60
           MK++F  +E APFAK+GGL DV GALPK L + G D  V++P Y+ +  ++  +++ +
Sbjct: 1   MKVLFAVSECAPFAKSGGLADVAGALPKELRRLGIDARVMLPKYETIAPEWKKKMKKVAE 60

Query: 61  FYTDVGWRHQYVGVKRLSQDNVTFYFIDNQYYFYRGHVYGDWDDGERFAYFQLAALELME 120
               VGWR QY GV+ L  D V +YFIDN+YYF R  +YG +DDGERFAYF  A LE++
Sbjct: 61  LIVPVGWRRQYCGVEELRHDGVIYYFIDNEYYFKRPQLYGHYDDGERFAYFCRAVLEVLP 120

Query: 121 KIDFIPDVLHVHDYHTAMIPFLLKEKYHWIQAYNNIRAVFTIHNIEFQGQFGPEMLGDLF 180
           +I F PDV+H HD+HT M+PFLL+E+Y     Y ++R VFTIHN++FQG F   +L DL
Sbjct: 121 EIQFQPDVIHCHDWHTGMVPFLLREQYRHELFYVDMRTVFTIHNLQFQGLFPRGILEDLL 180

Query: 181 GVGAERYEDGTLRWNNCLNWMKAAILYSDRVTTVSPSYANEIKTPEFGKGLDQIMRMEAG 240
            +    +    L +  C+++MK A++ SD +TTVSP+Y  EI+T  +G+ LD ++R
Sbjct: 181 NLDGRYFTVDHLEFYGCVSFMKGALVASDLITTVSPTYKEEIQTAYYGERLDGLLRARRD 240

Query: 241 KLSGIVNGIDSDLLNPETDAFLPYHFSKSNLEGKIKNKLALQENLGLPQDKNVPLIGIVS 300
            L GI+NGID +  NPE D FL   +S    E K  NK ALQ   GLP+  +VPLI +V+
Sbjct: 241 DLLGILNGIDDEFYNPEADPFLTATYSVHTRERKQLNKRALQRQFGLPEWDDVPLIAMVT 300
```

```
Query: 301 RLTDQKGFDIIASELDNMLQQDIQMVILGTGYHHFEETFSYFASRYPEKLSANITFDLRL 360
           R+T QKG D++      M+ +D+Q+V+LGTG   FE+ FS  A+ YP K+   I F   L
Sbjct: 301 RMTAQKGLDLVTCVFHEMMSEDMQLVVLGTGDWRFEQFFSQMAAAYPGKVGVYIGFHEPL 360

Query: 361 AQQIYAASDIFMMPSAFEPCGLSQMMAMRYGSLPLVHEVGGLKDTVVAFNQFDGSGTGFS 420
           A QIYA +D+F++PS FEPCGLSQM+A+RYG++P+V E GGL DTV ++N+    G GFS
Sbjct: 361 AHQIYAGADLFLIPSLFEPCGLSQMIALRYGTIPIVRETGGLNDTVQSYNEITKEGNGFS 420

Query: 421 FNHFSGYWLMQTLKLALEVYNDYPEAWKKLQWQAMSKDFSWDTACVAYEQLYQQL 475
           F +F+ + ++ T++ AL  Y   P  W++L  +AM  D+SW  +   Y+Q Y+QL
Sbjct: 421 FTNFNAHDMLYTIRRALSFYRQ-PSVWEQLTERAMRGDYSWRRSANQYKQAYEQL 474
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1420

A DNA sequence (GBSx1505) was identified in *S.agalactiae* <SEQ ID 4361> which encodes the amino acid sequence <SEQ ID 4362>. This protein is predicted to be a subunit of ADP-glucose pyrophosphorylase. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3492(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA19590 GB:D87026 subunit of ADP-glucose pyrophosphorylase
           [Bacillus stearothermophilus]
 Identities = 59/178 (33%), Positives = 111/178 (62%), Gaps = 1/178 (0%)

Query: 37  SAEIYVIDTPWLIEKMEEEAQNNEPRKLRFLLRDLIVESNALAFEYTGYLSNISSIKSYY 96
           S E+Y+++T  L++ + +  +N+    +  ++RD     +   +EY+GY + I S++ Y+
Sbjct: 157 SLEMYLLETSLLLDLIADY-KNHGYYSIVDVIRDYHRSLSICEYEYSGYAAVIDSVEQYF 215

Query: 97  DANMDMLTPNKFYSLFFSNQKVYTKVKNEEATYFDKQSNVSNSQLASGSIIKGYLDHSIV 156
            ++M++L  + +  LF  +  +YTKVK+E  T + ++ NV  S +A+G +I+G +++S++
Sbjct: 216 RSSMELLDRDVWEQLFLPSHPIYTKVKDEPPTKYGREGNVKRSMIANGCVIEGTVENSVL 275

Query: 157 SRNCLLEKGTRVVNSIIFPKVKIGEGATIENTIIDKCVKVASGVTLKGSLDKPLVIPK 214
            R+  + KG  V NSII  K +IG+G  ++  IIDK  KV  GV LKG+ ++P ++ K
Sbjct: 276 FRSVKIGKGAVVRNSIIMQKCQIGDGCVLDGVIIDKDAKVEPGVVLKGTKEQPFIVRK 333
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1421

A DNA sequence (GBSx1506) was identified in *S.agalactiae* <SEQ ID 4363> which encodes the amino acid sequence <SEQ ID 4364>. This protein is predicted to be subunit of ADP-glucose pyrophosphorylase (glgC-1). Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
```

```
----- Final Results -----
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9765> which encodes amino acid sequence <SEQ ID 9766> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA19589 GB:D87026 subunit of ADP-glucose pyrophosphorylase
          [Bacillus stearothermophilus]
 Identities = 195/352 (55%), Positives = 259/352 (73%)

Query: 7   MKNEMLALILAGGQGTRLGKLTQSIAKPAVQFGGRYRIIDFALSNCANSGINNVGVITQY 66
           MK + +A++LAGGQG+RL  LT +IAKPAV FGG+YRIIDF LSNC NSGI+ VGV+TQY
Sbjct: 1   MKKKCIAMLLAGGQGSRLRSLTTNIAKPAVPFGGKYRIIDFTLSNCTNSGIDTVGVLTQY 60

Query: 67  QPLELNTHIGNGSSWGLDGIDSGVTVLQPYSATEGNRWFQGTSHAIYQNIDYIDRINPEY 126
           QPL L+++IG GS+W LD  + GVTVL PYS + G +W++GT++A+YQNI+YI++ NP+Y
Sbjct: 61  QPLLLHSYIGIGSAWDLDRRNGGVTVLPPYSVSSGVKWYEGTANAVYQNINYIEQYNPDY 120

Query: 127 VLILSGDHIYKMNYDDMLQTHKDNLASLTVAVLDVPLKEASRFGIMNTDSNDRIVEFEEK 186
           VL+LSGDHIYKM+Y  ML  H    A +T++V++VP +EASRFGIMNT+    IVEF EK
Sbjct: 121 VLVLSGDHIYKMDYQHMLDYHIAKQADVTISVIEVPWEEASRFGIMNTNEEMEIVEFAEK 180

Query: 187 PEHPKSTKASMGIYIFDWKRLRTVLIDGEKNGIDMSDFGKNVIPAYLESGERVYTYNFDG 246
           P  PKS  ASMGIYIF+W  L+  L     N     DFGK+VIP  L   +R + Y F+G
Sbjct: 181 PAEPKSNLASMGIYIFNWPLLKQYLQIDNANPHSSHDFGKDVIPMLLREKKRPFAYPFEG 240

Query: 247 YWKDVGTIESLWEANMEYIGEDNKLHSRDRSWKIYSKNLIAPPNFMTEDANVKDSLVVDG 306
           YWKDVGT++SLWEANM+ + E+N+L   DRSW+IYS N   PP +++ +A V DSLV +G
Sbjct: 241 YWKDVGTVKSLWEANMDLLDENNELDLFDRSWRIYSVNPNQPPQYISPEAEVSDSLVNEG 300

Query: 307 CFVAGNVEHSILSTNVQVKPNAIIKDSFVMSGATIGEGAKINRAIIGEDAVI 358
           C V G VE S+L   V++   A++K+S +M GA + EGA + RAI+  D++I
Sbjct: 301 CVVEGTVERSVLFQGVRIGKGAVVKESVIMPGAAVSEGAYVERAIVTPDSII 352
```

There is also homology to SEQ ID 2660.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1422

A DNA sequence (GBSx1507) was identified in *S.agalactiae* <SEQ ID 4365> which encodes the amino acid sequence <SEQ ID 4366>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2844(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA78440 GB:Z14057 1,4-alpha-glucan branching enzyme [Bacillus
          caldolyticus]
 Identities = 272/616 (44%), Positives = 371/616 (60%), Gaps = 14/616 (2%)

Query: 6   ELYTFGIGENFHLQNYLGVHSENGSFC----FRVWAPNAENVQVIGDFTDWRNRPLQMNK 61
           E+Y F  G  +      G H   G       F VWAP+A  V+++G F DW      + K
```

```
Sbjct: 10  EVYLFHEGRLYQSYELFGAHVIRGGGAVGTRFCVWAPHAREVRLVGSFNDWNGTNSPLTK 69

Query: 62  -NQAGVWEANSLDAREGDLYKYLVTRKGGQVVEKIDPMAVYMERRPGTASVIKVLRNKKW 120
            N  GVW     +   EG LYKY +    G+V+ K DP A Y E RP TAS++  L+  +W
Sbjct: 70  VNDEGVWTIVVPENLEGHLYKYEIITPDGRVLLKADPYAFYSELRPHTASIVYDLKGYEW 129

Query: 121 EDGLWMGRRKRLGFQKRPINIYEVHAGSWKKDDFGHPMTFSQLKDYLIPYLVEMNYTHVE 180
            D  W  +++R     +P+ IYE+H GSWKK   G   T+ ++ D LIPY++E  +TH+E
Sbjct: 130 NDSPWQRKKRRKRIYDQPMVIYELHFGSWKKKPDGRFYTYREMADELIPYVLERGFTHIE 189

Query: 181 FMPLMAHPLDMSWGYQLMGYFAFEHTYGTPEEFQDFVEACHKNNIGVLVDWVPGHFIQND 240
            +PL+ HPLD SWGYQ  GY++    YGTP +F  FV+ CH+  +GV++DWVPGHF ++
Sbjct: 190 LLPLVEHPLDRSWGYQGTGYYSVTSRYGTPHDFMYFVDRCHQAGLGVIIDWVPGHFCKDA 249

Query: 241 DALAYFDGTATYEYQNHDRAHNYRWGALNFDLGKNQVQSFLISSALFWIEHYHIDGIRVD 300
             L  FDG  TYEY N     NY WG  NFDLGK +V+SFLIS+ALFW+E+YH+DG RVD
Sbjct: 250 HGLYMFDGAPTYEYANEKDRENYVWGTANFDLGKPEVRSFLISNALFWLEYYHVDGFRVD 309

Query: 301 AVSNMLYLDYDEGPWEANQFGDNRNLEGYHFLRKLNKVIKERHPNVMMIAEESTASTPIT 360
           AV+NMLY   ++  +E        N     FLR+LN+ +    PNV MIAE+ST    +T
Sbjct: 310 AVANMLYWPNNDRLYE--------NPYAVEFLRQLNEAVFAYDPNVWMIAEDSTDWPRVT 361

Query: 361 KDLESGGLGFDFKWNMGWMNDILRFYEEDPLYRQYDFNLVTFSFMYIFNENFVLAFSHDE 420
                GGLGF++KWNMGWMND+L++ E  P  R+Y  N V+FS +Y ++ENF+L FSHDE
Sbjct: 362 APTYDGGLGFNYKWNMGWMNDMLKYMETPPHERKYAHNQVSFSLLYAYSENFILPFSHDE 421

Query: 421 VVHGKKSMMHKMWGDRYNQFAGLRNLYAYQMCHPGKKLLFMGSEFGQFLEWKYNDQLEWE 480
           VVHGKKS+++KM G    +FA LR LY Y M HPGKKLLFMGSEF QF EWK+ ++L+W
Sbjct: 422 VVHGKKSLLNKMPGSYEEKFAQLRLLYGYMMAHPGKKLLFMGSEFAQFDEWKFAEELDWV 481

Query: 481 NLNDDMNQKMQRYTKQLNQFYKDHKCLWRIDDSFDGLEIIDADNKSETVLSFIRKDDK-G 539
             + ++++KM  Y KQL   YK +K  + +D    G E ID  N  +++ SFIR+  K G
Sbjct: 482 LFDFELHRKMDEYVKQLIACYKRYKPFYELDHDPRGFEWIDVHNAEQSIFSFIRRGKKEG 541

Query: 540 DLLLCVFNMTPVERPNFTIGVPQAGIYEEVLNTEMEEFGGVWKNHNPVTKTQVATWKDYD 599
           D+L+ V N T     ++ + VP    Y EVLN++  EFGG    +          +
Sbjct: 542 DVLVIVCNFTNQAYDDYKVSVPLLAPYREVLNSDAAEFGGSGHVNGKRLPAFSEPFHGKP 601

Query: 600 HTLSFTLPALGASVWR 615
           + +  T+P  G S+ R
Sbjct: 602 YHVRMTIPPFGISILR 617
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1423

A DNA sequence (GBSx1508) was identified in *S.agalactiae* <SEQ ID 4367> which encodes the amino acid sequence <SEQ ID 4368>. This protein is predicted to be pullulanase (pulA). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3194(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44685 GB:U67061 pullulanase [Bacteroides thetaiotaomicron]
 Identities = 223/597 (37%), Positives = 331/597 (55%), Gaps = 55/597 (9%)

Query: 139 EYSETKTAFRLWAPTAERVELILYHSTDETASVSKVLSMKRGTAVNYKNHKENTHGVWFT 198
```

```
           EY+   T F LW+PTA+ V L+LY +  E     + + M+ G             G W
Sbjct: 46  EYTPEATKFTLWSPTADEVRLMLYEA-GEGGHAYETVKMQSGE-----------EGTWTA 93

Query: 199 ELEGNYNYQAYTYRVYYRRRTFKITRDPYSIATTANGKRSIVIAPEALTPEGFKISHGKE 258
            +  +   + YT+ V    +    T    + A  NGKR+ +I  ++   P+G++     +
Sbjct: 94  VVSKDLIGKFYTFNVKIDDKWQGDTPGINARAVGVNGKRAAIIDWQSTNPDGWE----SD 149

Query: 259 AKWRLENPNQAVIYEMHVRDFSISETSGVKTDYHGKFKGLHQKGTVNQHGDKTTFDYVQD 318
            +  L++P   +IYEMH RDFS+  TSGVK    GK+  L + GT+N     T  D++ +
Sbjct: 150 TRPPLKSPADMIIYEMHHRDFSVDSTSGVKNK--GKYLALTEHGTMNSDKLLTGIDHLIE 207

Query: 319 LGVNYIQLQPIFDHHQTFDDD-GHYAYNWGYDPENYNVPEASFSSNPHEPATRILELKSA 377
           LGV ++ L P FD+    +      +YNWGYDP+NYNVP+ S++++P++PATR+ E K
Sbjct: 208 LGVTHVHLLPSFDYASVDETRLNENSYNWGYDPQNYNVPDGSYATDPYQPATRVKEFKQM 267

Query: 378 IQAYHDAGIGVIMDVVYNHTFSSTDSAFQLTVPDYYYRMNHNGTFQNGSGCGNETASEKE 437
           +QA H AGI VIMDVVYNHTF++ +S F+ TVP Y+YR   + T  NGSGCGNETASE+
Sbjct: 268 VQALHKAGIRVIMDVVYNHTFNTDESNFERTVPGYFYRQKEDKTLANGSGCGNETASERL 327

Query: 438 MCRKYILDSVLYWVKEYNIDGFRFDLMGLHDVETMNIIRNELNKIDPRILVYGEGWDMGA 497
           M RK++++SVLYW+KEY++DGFRFDLMG+HD+ETMN IR  +N +DP I +YGEGW   A
Sbjct: 328 MMRKFMVESVLYWIKEYHVDGFRFDLMGIHDIETMNEIRKAVNAVDPTICIYGEGWAAEA 387

Query: 498 GLTPQNK-AKKDNAYQMPGIGFFNDDVRDAV---KGAEIYGEFKKGLVSGNSTEDIVAKG 553
              P +  A K N  Q+PG+  F+D++RD +    G +  G F  G+  G   E  V  G
Sbjct: 388 PQYPADSLAMKGNIAQIPGVAVFSDELRDGLCGPVGDKRKGAFLAGIPGG---EMSVKFG 444

Query: 554 ILGSDE-------LVSYI------DPSQVLNYVEAHDNYNLNDLLWELHPNDNEKQHIYR 600
           I G+ E        V+Y        P Q+++YV  HD   L D L    P+   +Q I
Sbjct: 445 IAGAIEHPQVQCDSVNYTQKPWAKQPVQMISYVSCHDGLCLVDRLKASMPDITPEQLIRL 504

Query: 601 VEVASAMNLLMQGMAFMQLGQEFLRTKCYPTGDKGQLTQADKERAMNSYNAPDQVNQVNW 660
            ++A A+    QG+ F+  G+E +R               DK+   NSY +PD VN ++W
Sbjct: 505 DKLAQAVVFTSQGIPFIYAGEEIMR---------------DKQGVDNSYKSPDAVNAIDW 549

Query: 661 DNVTFHKSTINFIRKIITLKTNSPYFSYSSFEEIRKHVFVESAQYHSGFISFTVEEH 717
              T       + +++I L+ + P F      ++RKH+     +  S  I+F +++H
Sbjct: 550 RRKTTSADVFMYYKRLIDLRKSHPAFRMGDAGQVRKHLEFLPVE-GSNLIAFRLKDH 605
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1424

A DNA sequence (GBSx1509) was identified in *S.agalactiae* <SEQ ID 4369> which encodes the amino acid sequence <SEQ ID 4370>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2368(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 151/293 (51%), Positives = 193/293 (65%), Gaps = 5/293 (1%)

Query: 5   KKARLIYNPTSGQEIMKKNVAEVLDILEGFGYETSAFQTTPTKNSARDEATRAAQAGFDL 64
           K+AR+IYNPTSG+EI KK++A+VL   E  GYETS   TT     A   A  AA   FDL
Sbjct: 2   KRARIIYNPTSGREIFKKHLAQVLQKFEQAGYETSTHATT-CAGDATHAAKEAALREFDL 60

Query: 65  IVAAGGDGTINEVVNGIAPLKRRPKMAIIPTGTTNDFARALKIPRGNPIEATKLIGKNQI 124
           I+AAGGDGTINEVVNG+APL  RP + +IP GTTNDFARAL IPR + ++A   +
```

```
Sbjct: 61  IIAAGGDGTINEVVNGLAPLDNRPTLGVIPVGTTNDFARALGIPREDILKAADTVINGVA 120

Query: 125 VKMDIGQAQEDNYFINIAAAGSLTELTYSVPSQLKTTFGYLAYLAKGVELLPRVRKVPVK 184
             +DIGQ     YFINIA  G LTELTY VPS+LKT  G LAY  KG+E+LP +R   V+
Sbjct: 121 RPIDIGQVN-GQYFINIAGGGRLTELTYDVPSKLKTMLGQLAYYLKGMEMLPSLRPTEVE 179

Query: 185 ITHDKGEFIGDASMIFVAITNSVGGFEQIAPDAKLDDGKFTLILVKTANLIEIMHLIRLV 244
           I +D   F G+  +  V +TNSVGGFE++APD+ L+DG F L+++K ANL E + +  +
Sbjct: 180 IEYDGKLFQGEIMLFLVTLTNSVGGFEKLAPDSSLNDGMFDLMILKKANLAEFIRVATMA 239

Query: 245 LAGGKHINDKRVEYIKTSYLTIEPLSDERMMINLDGEYGGDAPITLANLKNHI 297
           L  G+HIND+ + Y K + + +     E+M +NLDGEYGG  P    NL  HI
Sbjct: 240 LR-GEHINDQHIIYTKANRVKVN--VSEKMQLNLDGEYGGMLPGEFVNLYRHI 289
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4371> which encodes the amino acid sequence <SEQ ID 4372>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2501(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 272/334 (81%), Positives = 300/334 (89%)

Query: 1   MKKQKKARLIYNPTSGQEIMKKNVAEVLDILEGFGYETSAFQTTPTKNSARDEATRAAQA 60
           MKKQ +ARLIYNPTSGQE+M+K+V EVLDILEGFGYETSAFQTT  KNSA +EA RAA+A
Sbjct: 1   MKKQLRARLIYNPTSGQELMRKSVPEVLDILEGFGYETSAFQTTAQKNSALNEARRAAKA 60

Query: 61  GFDLIVAAGGDGTINEVVNGIAPLKRRPKMAIIPTGTTNDFARALKIPRGNPIEATKLIG 120
           GFDL++AAGGDGTINEVVNGIAPLK+RPKMAIIPTGTTNDFARALK+PRGNP +A KLIG
Sbjct: 61  GFDLLIAAGGDGTINEVVNGIAPLKKRPKMAIIPTGTTNDFARALKVPRGNPSQAAKLIG 120

Query: 121 KNQIVKMDIGQAQEDNYFINIAAAGSLTELTYSVPSQLKTTFGYLAYLAKGVELLPRVRK 180
           KNQ ++MDIG+A++D YFINIAAAGSLTELTYSVPSQLKT FGYLAYLAKGVELLPRV
Sbjct: 121 KNQTIQMDIGRAKKDTYFINIAAAGSLTELTYSVPSQLKTMFGYLAYLAKGVELLPRVSN 180

Query: 181 VPVKITHDKGEFIGDASMIFVAITNSVGGFEQIAPDAKLDDGKFTLILVKTANLIEIMHL 240
           VPVKITHDKG F G  SMIF AITNSVGGFE IAPDAKLDDG FTLIL+KTANL EI+HL
Sbjct: 181 VPVKITHDKGVFEGQVSMIFAAITNSVGGFEMIAPDAKLDDGMFTLILIKTANLFEIVHL 240

Query: 241 IRLVLAGGKHINDKRVEYIKTSYLTIEPLSDERMMINLDGEYGGDAPITLANLKNHIRFF 300
           +RL+L GGKHI D+RVEYIKTS + IEP   +RMMINLDGEYGGDAPITL NLKNHI FF
Sbjct: 241 LRLILDGGKHITDRRVEYIKTSKIVIEPQCGKRMMINLDGEYGGDAPITLENLKNHITFF 300

Query: 301 ANTDEISDDALVLDKDELAIEAIAQKFANEVDDL 334
           A+TD ISDDALVLD+DEL IE I +KFA+EV+DL
Sbjct: 301 ADTDLISDDALVLDQDELEIEEIVKKFAHEVEDL 334
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1425

A DNA sequence (GBSx1510) was identified in *S.agalactiae* <SEQ ID 4373> which encodes the amino acid sequence <SEQ ID 4374>. This protein is predicted to be DNA ligase (ligA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -0.27   Transmembrane  363 - 379 ( 363 - 379)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9763> which encodes amino acid sequence <SEQ ID 9764> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12482 GB:Z99107 similar to DNA ligase [Bacillus subtilis]
 Identities = 346/657 (52%), Positives = 462/657 (69%), Gaps = 8/657 (1%)

Query: 2   ENRMNELVSLLNQYAKEYYTQDNPTVSDSQYDQLYRELVELEKQHPENILPNSPTHRVGG 61
           + R  EL   +N+Y+ EYYT D P+V D++YD+L +EL+ +E++HP+   P+SPT RVGG
Sbjct: 7   KQRAEELRRTINKYSYEYYTLDEPSVPDAEYDRLMQELIAIEEEHPDLRTPDSPTQRVGG 66

Query: 62  LVLEGFEKYQHEYPLYSLQDAFSKEELIAFDKRVKAEF-PTAAYMAELKIDGLSVSLTYV 120
            VLE F+K  H  P+ SL +AF+ ++L  FD+RV+       AY  ELKIDGL+VSL Y
Sbjct: 67  AVLEAFQKVTHGTPMLSLGNAFNADDLRDFDRRVRQSVGDDVAYNVELKIDGLAVSLRYE 126

Query: 121 NGVLQVGATRGDGNIGENITENLKRVHDIPLHLDQSLDITVRGECYLPKESFEAINIEKR 180
           +G    GATRGDG  GE+ITENLK + +IPL +++ L I VRGE Y+PK SFEA+N E+
Sbjct: 127 DGYFVRGATRGDGTTGEDITENLKTIRNIPLKMNRELSIEVRGEAYMPKRSFEALNEERI 186

Query: 181 ANGEQEFANPRNAAAGTLRQLNTGIVAKRKLATFLYQEASPTQK--ETQDDVLKELESYG 238
            N E+ FANPRNAAAG+LRQL+  I AKR L  F+Y  A   +   ETQ   L  L+  G
Sbjct: 187 KNEEEPFANPRNAAAGSLRQLDPKIAAKRNLDIFVYSIAELDEMGVETQSQGLDFLDELG 246

Query: 239 FSVNHHRLISSSMEKIWDFIQTIEKDRVSLPYDIDGIVIKVNSIAMQEELGFTVKAPRWA 298
           F  N  R    S+E++   I  ++  R  LPY+IDGIVIKV+S+  QEELGFT K+PRWA
Sbjct: 247 FKTNQERKKCGSIEEVITLIDELQAKRADLPYEIDGIVIKVDSLDQQEELGFTAKSPRWA 306

Query: 299 IAYKFPAEEKEAEILSVDWTVGRTGVVTPTANLTPVQLAGTTVSRATLHNVDYIAEKDIR 358
           IAYKFPAEE   ++L ++  VGRTGV+TPTA L PV++AGTTVSRA+LHN D I EKDIR
Sbjct: 307 IAYKFPAEEVVTKLLDIELNVGRTGVITPTAILEPVKVAGTTVSRASLHNEDLIKEKDIR 366

Query: 359 IGDTVVVYKAGDIIPAVLNVVMSKRNQQEVML-IPKLCPSCGSELVHFEGEVALRCINPL 417
           I D VVV KAGDIIP V+NV++ +R  +E    +P  CP CGSELV  EGEVALRCINP
Sbjct: 367 ILDKVVVKKAGDIIPEVVNVLVDQRTGEEKEFSMPTECPECGSELVRIEGEVALRCINPE 426

Query: 418 CPNQIKERLAHFASRDAMNITGFGPSLVEKLFDAHLIADVADIYRLSIENLLTLDGIKEK 477
           CP QI+E L HF SR+AMNI G G  ++ +LF+ +L+ +VAD+Y+L+ E ++ L+ + EK
Sbjct: 427 CPAQIREGLIHFVSRNAMNIDGLGERVITQLFEENLVRNVADLYKLTKERVIQLERMGEK 486

Query: 478 SATKIYHAIQSSKENSAEKLLFGLGIRHVGSKASRLLLEEFGNLRQLSQASQESIASIDG 537
           S   +  +IQ SKENS E+LLFGLGIR +GSKA++ L   F +L  L +AS+E + ++D
Sbjct: 487 STENLISSIQKSKENSLERLLFGLGIRFIGSKAAKTLAMHFESLENLKKASKEELLAVDE 546

Query: 538 LGGVIAKSLHTFFEKEEVDKLLEELTSYNVNFNYLG----KRVSTDAQLSGLTVVLTGKL 593
           +G  +A ++ T+F KEE+ +LL EL    VN  Y G    K   +D+  +G T+VLTGKL
Sbjct: 547 IGEKMADAVITYFHKEEMLELLNELQELGVNTLYKGPKKVKAEDSDSYFAGKTIVLTGKL 606

Query: 594 EKMTRNEAKEKLQNLGAKVTGSVSKKTDLIVAGSDAGSKLTKAQDLGITIQDEDWLL 650
           E+++RNEAK +++ LG K+TGSVSK TDL++AG  AGSKLTKAQ+L I + +E+ L+
Sbjct: 607 EELSRNEAKAQIEALGGKLTGSVSKNTDLVIAGEAAGSKLTKAQELNIEVWNEEQLM 663
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4375> which encodes the amino acid sequence <SEQ ID 4376>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43   Transmembrane  363 - 379 ( 363 - 379)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 472/652 (72%), Positives = 556/652 (84%)

Query: 1   MENRMNELVSLLNQYAKEYYTQDNPTVSDSQYDQLYRELVELEKQHPENILPNSPTHRVG 60
           M+ R+ EL  LLN+Y  +YYT+D P+VSDS YD+LYRELV LE+ +PE +L +SPT +VG
Sbjct: 1   MKKRIKELTDLLNRYRYDYYTKDAPSVSDSDYDKLYRELVTLEQSYPEYVLQDSPTQQVG 60

Query: 61  GLVLEGFEKYQHEYPLYSLQDAFSKEELIAFDKRVKAEFPTAAYMAELKIDGLSVSLTYV 120
           G +L+GFEKY+H+YPL+SLQDAFS+EEL AFDKRVKAEFP A Y+AELKIDGLS+SL+Y
Sbjct: 61  GTILKGFEKYRHQYPLFSLQDAFSREELDAFDKRVKAEFPNATYLAELKIDGLSISLSYE 120

Query: 121 NGVLQVGATRGDGNIGENITENLKRVHDIPLHLDQSLDITVRGECYLPKESFEAINIEKR 180
           NG LQVGATRGDGNIGENITEN+K++ DIP  L + L ITVRGE Y+ ++SF+AIN  ++
Sbjct: 121 NGFLQVGATRGDGNIGENITENIKKIKDIPYQLSEPLTITVRGEAYMSRQSFKAINEARQ 180

Query: 181 ANGEQEFANPRNAAAGTLRQLNTGIVAKRKLATFLYQEASPTQKETQDDVLKELESYGFS 240
            NGE EFANPRNAAAGTLRQL+T +VAKR+LATFLYQEASPT +  Q++VL EL   GFS
Sbjct: 181 ENGETEFANPRNAAAGTLRQLDTSVVAKRQLATFLYQEASPTARNQQNEVLAELADLGFS 240

Query: 241 VNHHRLISSSMEKIWDFIQTIEKDRVSLPYDIDGIVIKVNSIAMQEELGFTVKAPRWAIA 300
           VN +  ++SSM++IWDFI+TIE  R  L YDIDG+VIKVNS+AMQEELGFTVKAPRWAIA
Sbjct: 241 VNPYYQLTSSMDEIWDFIKTIEAKRDQLAYDIDGVVIKVNSLAMQEELGFTVKAPRWAIA 300

Query: 301 YKFPAEEKEAEILSVDWTVGRTGVVTPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG 360
           YKFPAEEKEAEILSVDWTVGRTGVVTPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG
Sbjct: 301 YKFPAEEKEAEILSVDWTVGRTGVVTPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG 360

Query: 361 DTVVVYKAGDIIPAVLNVVMSKRNQQEVMLIPKLCPSCGSELVHFEGEVALRCINPLCPN 420
           DTV+VYKAGDIIPAVLNVVMSKRNQQEVMLIPKLCPSCGSELVHFE EVALRCINPLCP+
Sbjct: 361 DTVIVYKAGDIIPAVLNVVMSKRNQQEVMLIPKLCPSCGSELVHFEDEVALRCINPLCPS 420

Query: 421 QIKERLAHFASRDAMNITGFGPSLVEKLFDAHLIADVADIYRLSIENLLTLDGIKEKSAT 480
            I+  L HFASRDAMNITG GP++VEKLF A  + DVADIY+L+ E+ + LDGIKEKSA
Sbjct: 421 LIQRSLEHFASRDAMNITGLGPAIVEKLFLAGFVHDVADIYQLTKEDFMQLDGIKEKSAD 480

Query: 481 KIYHAIQSSKENSAEKLLFGLGIRHVGSKASRLLLEEFGNLRQLSQASQESIASIDGLGG 540
           K+  AI++SK NSAEKLLFGLGIRH+GSK SRL+LE +G++  L  A +E IA IDGLG
Sbjct: 481 KLLAAIEASKSNSAEKLLFGLGIRHIGSKVSRLILEVYGDISALLTAKEEEIARIDGLGS 540

Query: 541 VIAKSLHTFFEKEEVDKLLEELTSYNVNFNYLGKRVSTDAQLSGLTVVLTGKLEKMTRNE 600
            IA+SL  +FE++    L++EL +  VN +Y G++V++DA L GLTVVLTGKL ++ RNE
Sbjct: 541 TIAQSLTQYFEQKTAAILVDELKTAGVNMHYSGQKVNSDAALFGLTVVLTGKLNQLNRNE 600

Query: 601 AKEKLQNLGAKVTGSVSKKTDLIVAGSDAGSKLTKAQDLGITIQDEDWLLNL 652
           AK+KL+ LGAKVTGSVSKKTDL++AGSDAGSKL KA+ LGI I+DEDWL  L
Sbjct: 601 AKDKLEALGAKVTGSVSKKTDLVIAGSDAGSKLEKAKSLGIRIEDEDWLRQL 652
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1426

A DNA sequence (GBSx1511) was identified in *S.agalactiae* <SEQ ID 4377> which encodes the amino acid sequence <SEQ ID 4378>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.63    Transmembrane  110 - 126 ( 108 - 128)
   INTEGRAL    Likelihood = -2.13    Transmembrane  142 - 158 ( 141 - 159)
   INTEGRAL    Likelihood = -1.12    Transmembrane   75 -  91 (  75 -  93)
```

```
----- Final Results -----
                  bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA68244 GB:X99978 citrulline cluster-linked gene [Lactobacillus
          plantarum]
 Identities = 56/158 (35%), Positives = 91/158 (57%), Gaps = 8/158 (5%)

Query: 13  AIVTAIYIVLTITPPFNAIAYGAYQFRVSEMLNFLAFYHRKYLFAVTLGCMISNLYSFG- 71
            A+V A+Y+VL + P   ++A GA QFRVSE LN LA ++RKY++ +  G ++ + +  G
Sbjct: 13  ALVAAMYVVLCLGPAAFSLASGAIQFRVSEGLNHLAVFNRKYIWGIVAGVILFDAFGPGA 72

Query: 72  -MIDVFVGGGSTLLFVYLGTILFKQYQKDYLFNGLINKAFFFFSFFFAASMITVAVELKI 130
             +++V  GGG +LL + + T L  +  K      L+N A F  S F  A MIT+     +
Sbjct: 73  SLLNVLFGGGQSLLALLVLTWLAPKL-KTVWQRMLLNIALFTVSMFMIALMITM-----M 126

Query: 131 VAGLPLLLTWLTTAVGELASLLVGAVLVDKLSRHVDFT 168
             +G+    T+LTTA+ EL  + + A ++  L R + F+
Sbjct: 127 SSGVAFWPTYLTTALSELIIMSITAPIMYSLDRVLHFS 164
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4379> which encodes the amino acid sequence <SEQ ID 4380>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.41    Transmembrane    75 -  91 (  70 -  94)
   INTEGRAL    Likelihood = -3.82    Transmembrane    12 -  28 (   8 -  28)
   INTEGRAL    Likelihood = -2.28    Transmembrane   141 - 157 ( 140 - 158)
   INTEGRAL    Likelihood = -0.64    Transmembrane   110 - 126 ( 110 - 126)
   INTEGRAL    Likelihood = -0.59    Transmembrane    55 -  71 (  54 -  73)

----- Final Results -----
                  bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/167 (68%), Positives = 137/167 (81%), Gaps = 1/167 (0%)

Query: 1   MNTFTTRDYAHMAIVTAIYIVLTITPPFNAIAYGAYQFRVSEMLNFLAFYHRKYLFAVTL 60
           M   T  DY H+ +V A+Y+VLTITPP NAI+YG YQFR+SEM+NFLAFYHRKY+ AVTL
Sbjct: 1   MTKLTVHDYVHIGLVAALYVVLTITPPLNAISYGMYQFRISEMMNFLAFYHRKYIIAVTL 60

Query: 61  GCMISNLYSFGMIDVFVGGGSTLLFVYLGTILFKQYQKDYLFNGLINKAFFFFSFFFAAS 120
           GCMI+N YSFG+IDVFVGGGSTL+FV LG ILF +YQKDYLFNG+ NKAF +FSFFFA S
Sbjct: 61  GCMIANFYSFGLIDVFVGGGSTLIFVTLGVILFSKYQKDYLFNGIFNKAFVYFSFFFATS 120

Query: 121 MITVAVELKIVAGLPLLLTWLTTAVGELASLLVGAVLVDKLSRHVDF 167
           M  VA+EL    G P LLTW TTA+GEL SLL+G++++DKLS+ + F
Sbjct: 121 MFNVAIELYFF-GAPFLLTWFTTALGELVSLLIGSLIIDKLSQRISF 166
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1427

A DNA sequence (GBSx1513) was identified in *S.agalactiae* <SEQ ID 4381> which encodes the amino acid sequence <SEQ ID 4382>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.20   Transmembrane  255 - 271 ( 245 - 281)
   INTEGRAL    Likelihood =-10.72   Transmembrane  141 - 157 ( 132 - 165)
   INTEGRAL    Likelihood = -8.17   Transmembrane  189 - 205 ( 185 - 208)
   INTEGRAL    Likelihood = -7.01   Transmembrane   36 -  52 (  33 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.5479(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC35915 GB:AF071085 Orfde2 [Enterococcus faecalis]
 Identities = 83/276 (30%), Positives = 157/276 (56%), Gaps = 3/276 (1%)

Query: 17  RPIQVFMRHFQSAEMDLSAIAVAYYLLVTAFPLLVIAANIFPYFHINVSDLLSLMQKNLP 76
            R I+    H  +AE+  S++ VAYYLL++ FPLL+   N+ PY  I+ + +L  + + +P
Sbjct: 15  RFIETTQSHMVTAEIGNSSVVVAYYLLLSLFPLLIAVGNVLPYLRIDPNSVLPYIAEAIP 74

Query: 77  KNIYEPASRLAVDAFSKPSTGILGFASLTAFWTMSKSLTSLQKAINKAYGVDQHRDFVIS 136
            K++Y+          ++ S G+L  ++L AFW+ S+S+ +LQ A+NKA+GV+Q ++F++
Sbjct: 75  KDVYKNLEPAIRSLLTQRSGGLLSVSALAAFWSASQSINALQNAMNKAFGVEQRKNFILV 134

Query: 137 RLVGVGTGLIILFLLTFVLIFSTFSKPVLQIIVNMYDLGDTLTAWLLNLAQPVTFLTIFL 196
            R+V     L+ +  +  V++     + +++++  ++    ++      L  P+T + + +
Sbjct: 135 RVVSFLVILLFMVAIVGVVVILGLGQYIIELLQPIFHYSTSVIDTFQALKWPLTTVVLLV 194

Query: 197 GIGILYFILPNARIRKVRYVIPGTLFSTFVIGFFSNLISQYVLNRVEKMVDIKTFGSVVI 256
             + ++Y ++PN ++  +R ++PG +FST      S +   YV     ++   +  GS +
Sbjct: 195 IMCLIYAVVPNRKL-SLRSILPGAIFSTVGWMLLSQIFGLYVKYFSSRIASYQIIGSFI- 252

Query: 257 FILMLWFIFLAHIMILGAILNASVQEIATGKIESRR 292
             ILMLW  F A I+ILGAI+NA V E   G  E ++
Sbjct: 253 -ILMLWLNFAATIIILGAIVNAVVDEYLXGXKEKKQ 287
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4383> which encodes the amino acid sequence <SEQ ID 4384>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.58   Transmembrane  141 - 157 ( 132 - 168)
   INTEGRAL    Likelihood =-12.15   Transmembrane  189 - 205 ( 177 - 210)
   INTEGRAL    Likelihood =-11.68   Transmembrane  256 - 272 ( 245 - 280)
   INTEGRAL    Likelihood = -7.54   Transmembrane   36 -  52 (  33 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA68244 GB:X99978 citrulline cluster-linked gene [Lactobacillus
           plantarum]
 Identities = 53/170 (31%), Positives = 92/170 (53%), Gaps = 11/170 (6%)

Query: 1   MTKLTVHDYVHIGLVAALYVVLTITPPLNAISYGMYQFRISEMMNFLAFYHRKYIIAVTL 60
            MT+  +  ++   LVAA+YVVL + P   +++ G  QFR+SE +N LA ++RKYI  +
Sbjct: 1   MTQSKIRPWIINALVAAMYVVLCLGPAAFSLASGAIQFRVSEGLNHLAVFNRKYIWGIVA 60

Query: 61  GCMIANFYSFG--LIDVFVGGGSTLIFVTLGVILFSKYQKDYLFNGIFNKAFVYFSFFFA 118
            G ++ + +  G  L++V  GGG +L+ + +   L  K +       ++ +  +  +  F
Sbjct: 61  GVILFDAFGPGASLLNVLFGGGQSLLALLVLTWLAPKLKT------VWQRMLLNIA-LFT 113

Query: 119 TSMFNVA--IELYFFGAPFLLTWFTTALGELVSLLIGSLIIDKLSQRISF 166
```

```
            SMF +A  I +   G  F  T+ TTAL EL+ + I + I+  L + + F
Sbjct: 114 VSMFMIALMITMMSSGVAFWPTYLTTALSELIIMSITAPIMYSLDRVLHF 163
!GB:AF071085 Orfde2 [Enterococcus faecalis]         176  2e-43

>GP:AAC35915 GB:AF071085 Orfde2 [Enterococcus faecalis]
 Identities = 90/271 (33%), Positives = 155/271 (56%), Gaps = 3/271 (1%)

Query: 19  IQVFMRHLQSAEMDLSAIAVAYYLILTAFPLIVIAANIFPYLNIDIADLLRLMKQNLPKD 78
            I+    H+ +AE+  S++ VAYYL+L+ FPL++   N+ PYL ID   +L  + + +PKD
Sbjct: 17  IETTQSHMVTAEIGNSSVVVAYYLLLSLFPLLIAVGNVLPYLRIDPNSVLPYIAEAIPKD 76

Query: 79  IFRPASAIVENIFSKPSGSVLGVATLTGLWTMSRSLTSLQKAINKAYGASQHRDFFIGHL 138
            +++     + ++ ++ SG +L V+ L   W+ S+S+ +LQ A+NKA+G  Q ++F +  +
Sbjct: 77  VYKNLEPAIRSLLTQRSGGLLSVSALAAFWSASQSINALQNAMNKAFGVEQRKNFILVRV 136

Query: 139 VGLLTSLIILFLLAFALIFSIFSKAAIQVLDKHYHLSDNITTIFLLLIQPITVLIIFVGL 198
            V  L  L+ +  +   ++     +  I++L   +H S ++   F  L  P+T +++ V +
Sbjct: 137 VSFLVILLFMVAIVGVVVILGLGQYIIELLQPIFHYSTSVIDTFQALKWPLTTVVLLVIM 196

Query: 199 MLLYFLLPNVKIKKIRYILPGTLFTSFVMTFLSNLVGNYVVYNVERMVDIKMFGSVMIFI 258
             L+Y ++PN K+  +R ILPG +F++     LS + G YV Y   R+   ++ GS    I
Sbjct: 197 CLIYAVVPNRKL-SLRSILPGAIFSTVGWMLLSQIFGLYVKYFSSRIASYQIIGS--FII 253

Query: 259 IMLWFIFLARILILGAIFNATYQEMSLGKLE 289
            +MLW  F A I+ILGAI NA   E   G  E
Sbjct: 254 LMLWLNFAATIIILGAIVNAVVDEYLXGXKE 284
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 188/302 (62%), Positives = 244/302 (80%)

Query: 1   MKLKKFFEDLLAKLEYRPIQVFMRHFQSAEMDLSAIAVAYYLLVTAFPLLVIAANIFPYF 60
            M  KK+F+ +L+K +Y PIQVFMRH QSAEMDLSAIAVAYYL++TAFPL+VIAANIFPY
Sbjct: 1   MAEKKWFDKVLSKWQYEPIQVFMRHLQSAEMDLSAIAVAYYLILTAFPLIVIAANIFPYL 60

Query: 61  HINVSDLLSLMQKNLPKNIYEPASRLAVDAFSKPSTGILGFASLTAFWTMSKSLTSLQKA 120
            +I+++DLL LM++NLPK+I+ PAS +  + FSKPS  +LG A+LT  WTMS+SLTSLQKA
Sbjct: 61  NIDIADLLRLMKQNLPKDIFRPASAIVENIFSKPSGSVLGVATLTGLWTMSRSLTSLQKA 120

Query: 121 INKAYGVDQHRDFVISRLVGVGTGLIILFLLTFVLIFSTFSKPVLQIIVNMYDLGDTLTA 180
            INKAYG  QHRDF I  LVG+ T LIILFLL F LIFS FSK  +Q++   Y L D +T
Sbjct: 121 INKAYGASQHRDFFIGHLVGLLTSLIILFLLAFALIFSIFSKAAIQVLDKHYHLSDNITT 180

Query: 181 WLLNLAQPVTFLTIFLGIGILYFILPNARIRKVRYVIPGTLFSTFVIGFFSNLISQYVLN 240
              L L QP+T L IF+G+ +LYF+LPN +I+K+RY++PGTLF++FV+ F SNL+  YV+
Sbjct: 181 IFLLLIQPITVLIIFVGLMLLYFLLPNVKIKKIRYILPGTLFTSFVMTFLSNLVGNYVVY 240

Query: 241 RVEKMVDIKTFGSVVIFILMLWFIFLAHIMILGAILNASVQEIATGKIESRRGDIMSLIQ 300
             VE+MVDIK FGSV+IFI+MLWFIFLA I+ILGAI NA+ QE++ GK+E R GD++++++
Sbjct: 241 NVERMVDIKMFGSVMIFIIMLWFIFLARILILGAIFNATYQEMSLGKLEGRSGDMIAILK 300

Query: 301 KS 302
            K+
Sbjct: 301 KT 302
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1428

A DNA sequence (GBSx1514) was identified in *S.agalactiae* <SEQ ID 4385> which encodes the amino acid sequence <SEQ ID 4386>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4200(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1429

A DNA sequence (GBSx1515) was identified in *S.agalactiae* <SEQ ID 4387> which encodes the amino acid sequence <SEQ ID 4388>. This protein is predicted to be methionine aminopeptidase (map). Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2342(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9761> which encodes amino acid sequence <SEQ ID 9762> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC35914 GB:AF071085 methionine aminopeptidase A [Enterococcus
           faecalis]
 Identities = 101/207 (48%), Positives = 128/207 (61%), Gaps = 31/207 (14%)

Query: 1   MITLKSAREIEAMDRAGDFLASIHIGLRDIIKPGVDMWEVEEYVRRRCKEENVLPLQIGV 60
           MITLKS REIE MD +G+ LA +H  LR  IKPG+  W++E +VR   +    +  QIG
Sbjct: 1   MITLKSPREIEMMDESGELLADVHRHLRTFIKPGITSWDIEVFVRDFIESHGGVAAQIGY 60

Query: 61  DGAVMDYPYATCCGLNDEVAHAFPRHYTLKQGDLLKVDMVLSEPLDKSIVDVSSLNFDNV 120
           +G    Y YATCC +NDE+ H FPR   LK GDL+KVDM +
Sbjct: 61  EG----YKYATCCSINDEICHGFPRKKVLKDGDLIKVDMCVD------------------ 98

Query: 121 AQMKKYTETYSGGLADSCWAYAVGEVSQEVKDLMSVTREAMYIGIEKAVIGNRIGDIGAA 180
                      G ++DSCW+Y VGE + E+  LM VT++A+Y+GIE+A +GNRIGDIG A
Sbjct: 99  ---------LKGAISDSCWSYVVGESTPEIDRLMEVTKKALYLGIEQAQVGNRIGDIGHA 149

Query: 181 IQDYAESRGYGVVRDLVGHGVGPTMHE 207
           IQ Y E  GYGVVRD VGHG+GPT+HE
Sbjct: 150 IQTYVEGEGYGVVRDFVGHGIGPTIHE 176
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4389> which encodes the amino acid sequence <SEQ ID 4390>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2082(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 256/286 (89%), Positives = 273/286 (94%)

Query: 1   MITLKSAREIEAMDRAGDFLASIHIGLRDIIKPGVDMWEVEEYVRRRCKEENVLPLQIGV 60
           MITLKSAREIEAMDRAGDFLA IHIGLRDIIKPGVDMWEVE YVRRRCKE+NVLPLQIGV
Sbjct: 1   MITLKSAREIEAMDRAGDFLAGIHIGLRDIIKPGVDMWEVEAYVRRRCKEDNVLPLQIGV 60

Query: 61  DGAVMDYPYATCCGLNDEVAHAFPRHYTLKQGDLLKVDMVLSEPLDKSIVDVSSLNFDNV 120
           DG +MDYPYATCCGLNDEVAHAFPRHY LK+GDLLKVDMVLSEPLDKSIVDV++L+FDNV
Sbjct: 61  DGHMMDYPYATCCGLNDEVAHAFPRHYILKEGDLLKVDMVLSEPLDKSIVDVAALDFDNV 120

Query: 121 AQMKKYTETYSGGLADSCWAYAVGEVSQEVKDLMSVTREAMYIGIEKAVIGNRIGDIGAA 180
            +MKK+T +Y+GGLADSCWAYAVG  S E+K LM VT+EAMY GIEKAVIGNRIGDIGAA
Sbjct: 121 PEMKKWTGSYTGGLADSCWAYAVGTPSDEIKQLMDVTKEAMYRGIEKAVIGNRIGDIGAA 180

Query: 181 IQDYAESRGYGVVRDLVGHGVGPTMHEEPMVPNYGTAGRGLRLREGMVLTIEPMINTGTW 240
           +Q+YAES GYGVVRDLVGHGVGPTMHEEPMVPNYGTAGRGLRL+EGMVLT+EPMINTGTW
Sbjct: 181 VQEYAESFGYGVVRDLVGHGVGPTMHEEPMVPNYGTAGRGLRLKEGMVLTVEPMINTGTW 240

Query: 241 EIDTDMKTGWAHKTLDGGLSCQYEHQFVITKDGPVILTSQGEERTY 286
           EIDTD+KTGWAHKTLDGGLSCQYEHQFVITKDGPVILTSQGEERTY
Sbjct: 241 EIDTDIKTGWAHKTLDGGLSCQYEHQFVITKDGPVILTSQGEERTY 286
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1430

A DNA sequence (GBSx1516) was identified in *S.agalactiae* <SEQ ID 4391> which encodes the amino acid sequence <SEQ ID 4392>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3473(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9759> which encodes amino acid sequence <SEQ ID 9760> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06894 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 158/431 (36%), Positives = 270/431 (61%), Gaps = 6/431 (1%)

Query: 6   SKHQEILEYLENLAVGKRVSVRSISNHLKVSDGTAYRAIKEAENRGIVETRPRSGTVRVA 65
           +KH++IL+Y+ NL VG+++SVR I+  L+VS+GTAYRAIKEAEN+G+V T  R GT+R+
Sbjct: 3   TKHEQILQYITNLEVGEKISVRRIAKDLQVSEGTAYRAIKEAENQGLVSTIERVGTIRIE 62

Query: 66  QKAKVNIEKLTYAEIARISDSQVVAGIEGLSKEFSKFSIGAMTHRNIEKYLVQGGLLIVG 125
           +K K NIEKLTYAE+  I D QV+ G +GL K  ++F IGAM    + +Y+  G LLIVG
Sbjct: 63  KKQKENIEKLTYAEVVNIVDGQVLGGRDGLHKTLNRFVIGAMKLDAMMRYVEPGNLLIVG 122

Query: 126 DRDEIQHLALQHQNAILVTGGFNVSPSVCRLADKLQIPVMVTHYDTFTVSTMINHTLSNA 185
           +R ++  +AL+   A+L+TGGF+ S    +LAD+L +PV+ T YDTFTV+TMIN  + +
Sbjct: 123 NRYQVHQIALEAGAAVLITGGFDTSDEAIKLADELDLPVISTSYDTFTVATMINRAIYDQ 182

Query: 186 KIRTDLKTVEQVYQSQMDYGFLAQDDTVKEFNLLVKQTKNVRFPIVNQANVVVGVVSVQD 245
            I+ ++  V+ +     D  ++  ++ V +++ L ++T + R+P++++   + G+V+ +D
Sbjct: 183 LIKKEITLVDDILIPLQDTYYMTTENVVGKWHELNEKTGHSRYPVIDENMKIQGMVAAKD 242

Query: 246 ILGKDKEVKLATVMSKNIIVAKPRMSLANISQKMIFEDLNMMPVVSDDFELLGVITRRQA 305
           +L   +   +  VM+KN I    R S+A ++  M++E + ++PV+    +L+GV++R+
Sbjct: 243 VLNASRHTPIEKVMTKNPITVSERTSVAAVAHVMVWEGIELLPVIDSHRKLIGVVSRQDV 302
```

```
Query: 306 VENLSMSQ-----GTDLYTYSDQILSNLQIEDG-HFSFLVEPAMIDHTGSLTQGVLTEFL 359
           ++ L M Q     G  +       L+    + G  +   + P M +  G+++ GV+T  +
Sbjct: 303 LKALQMIQRQPHVGETIEDLMTNGLNESSSDQGDSYEVEITPQMTNQLGTISHGVMTSLV 362

Query: 360 KEICIRVLTRKHQRSIVVKQMTLYFLQPVQIDEIIMVTPTIISEKRREATLDLELKLENK 419
            E   RVL +  +  +VV+ +TLYFL+PVQID  + + P ++   R+   +D+E+  E +
Sbjct: 363 IESGSRVLRKYKKGDLVVENITLYFLKPVQIDSRLTIRPRVLEIGRKHGKIDVEMYHEGE 422

Query: 420 IIAKAMIAVKI 430
           I+ KA+   +I
Sbjct: 423 IVGKALFMAQI 433
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4393> which encodes the amino acid sequence <SEQ ID 4394>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3011(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 267/431 (61%), Positives = 351/431 (80%)

Query: 1   MIIVMSKHQEILEYLENLAVGKRVSVRSISNHLKVSDGTAYRAIKEAENRGIVETRPRSG 60
           +II+MSKHQ+IL+YLE LA+GK+VSVRSISNHLKVSDGTAYRAIKEAENRGIVET+PRSG
Sbjct: 1   VIIIMSKHQDILDYLEKLAIGKKVSVRSISNHLKVSDGTAYRAIKEAENRGIVETKPRSG 60

Query: 61  TVRVAQKAKVNIEKLTYAEIARISDSQVVAGIEGLSKEFSKFSIGAMTHRNIEKYLVQGG 120
           TVR+ +K +V I++LTY+EIARISDS+V+AG  GL  EFS+FSIGAMT +NI +YLV+GG
Sbjct: 61  TVRIEKKGRVRIDRLTYSEIARISDSEVLAGHAGLGHEFSRFSIGAMTQQNIRRYLVKGG 120

Query: 121 LLIVGDRDEIQHLALQHQNAILVTGGFNVSPSVCRLADKLQIPVMVTHYDTFTVSTMINH 180
           LLIVGDR+ IQ LAL++ NAILVTGGF VS  V  +A+  +IPVMVTHYDTFTV+TMINH
Sbjct: 121 LLIVGDRETIQLLALENHNAILVTGGFPVSKRVIEMANNQRIPVMVTHYDTFTVATMINH 180

Query: 181 TLSNAKIRTDLKTVEQVYQSQMDYGFLAQDDTVKEFNLLVKQTKNVRFPIVNQANVVVGV 240
            LSN +I+TDLKTVEQV     DYG+L +D +V+EFN L+K+T+ VRFP+++    V+GV
Sbjct: 181 ALSNIRIKTDLKTVEQVMIPITDYGYLCEDSSVEEFNTLIKKTRQVRFPVLDYKRKVIGV 240

Query: 241 VSVQDILGKDKEVKLATVMSKNIIVAKPRMSLANISQKMIFEDLNMMPVVSDDFELLGVI 300
           VS++D++ +    KL  VMSKN I A+P  SLANISQKMIFEDLNM+PV  ++  LLG+I
Sbjct: 241 VSMRDVVDQLPTTKLTKVMSKNPITARPNTSLANISQKMIFEDLNMLPVTDEENNLLGMI 300

Query: 301 TRRQAVENLSMSQGTDLYTYSDQILSNLQIEDGHFSFLVEPAMIDHTGSLTQGVLTEFLK 360
           TRRQA+ENL   Q  + YTYS+QILSNL+    ++  +VEP MID  G+++ GV++EFLK
Sbjct: 301 TRRQAMENLPNHQPNNPYTYSEQILSNLEETVDYYQVVVEPTMIDSAGNMSNGVISEFLK 360

Query: 361 EICIRVLTRKHQRSIVVKQMTLYFLQPVQIDEIIMVTPTIISEKRREATLDLELKLENKI 420
           EI IR LT+KHQ++I+++QM +YFL  +QI++ + + P II+E RR +T+D+E+ +++++
Sbjct: 361 EISIRALTKKHQKNIIIEQMMVYFLHAIQIEDELKIYPKIITENRRSSTIDIEIFVDDQV 420

Query: 421 IAKAMIAVKIN 431
           IAKA+I  KIN
Sbjct: 421 IAKAIITTKIN 431
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1431

A DNA sequence (GBSx1517) was identified in *S.agalactiae* <SEQ ID 4395> which encodes the amino acid sequence <SEQ ID 4396>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2837(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04556 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 56/185 (30%), Positives = 86/185 (46%), Gaps = 4/185 (2%)

Query: 7   MDIWTNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRTNLQFI-FPVQTSKKESDFLLV 65
           M+I    G    +ETE + LR     D  A +  AS     +++ +    S K+S+  L
Sbjct: 1   MEIEDIYGDLPTLETERLRLRKFYKDDAAAIYDYASNEQVTKYVLWETHQSIKDSEAFLA 60

Query: 66  HSFMK---EPLGVWAIEDKVSHKMFGVIRFENIDLSKKTAEIGYFLKESSWGQGIMTECL 122
            +  K   + +  WAIE K + +M G + F       KTAE+GY L E  WGQGIMTE +
Sbjct: 61  FALNKYDEKDVSPWAIELKRNERMIGTVDFVWWKPKDKTAELGYVLSEPYWGQGIMTEAV 120

Query: 123 KTLSFFAFREFGMDKLIIVTHKENIASQKVALKAHFKQSRSFKGSDRYTRRIRDYIEFQL 182
             L  F F    ++++      ENI+S +V  KA      + + +       RD+  + +
Sbjct: 121 NALVEFGFNNMELERIQAKCFAENISSARVMEKAGLIYEGTHRRAIYVKGAHRDFKVYAI 180

Query: 183 TRGDY 187
            R DY
Sbjct: 181 IREDY 185
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 667> which encodes the amino acid sequence <SEQ ID 668>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1096(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 94/177 (53%), Positives = 117/177 (65%)

Query: 7   MDIWTNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRTNLQFIFPVQTSKKESDFLLVH 66
           MDIWT L  FAF ET  V LRP  Y D   F+ + +   NL ++FP Q +K  SD+LLVH
Sbjct: 1   MDIWTKLAVFAFFETPKVILRPFRYEDHWDFYSMVNDTKNLYYVFPEQKTKAASDYLLVH 60

Query: 67  SFMKEPLGVWAIEDKVSHKMFGVIRFENIDLSKKTAEIGYFLKESSWGQGIMTECLKTLS 126
           SF+K PLG WAIEDK +H++ G IR E+ D   + A+IGYFL  + WGQGIMTE +  L
Sbjct: 61  SFIKFPLGQWAIEDKATHQVIGSIRIEHYDAKTRCADIGYFLNYAFWGQGIMTEVVIKLV 120

Query: 127 FFAFREFGMDKLIIVTHKENIASQKVALKAHFKQSRSFKGSDRYTRRIRDYIEFQLT 183
           + +F EFG+  L I+TH EN ASQKVA KA F+    FKGSDR T +I  Y  +QLT
Sbjct: 121 YLSFHEFGLKTLRIITHLENKASQKVAKKAGFQLKTCFKGSDRNTHKICIYKMYQLT 177
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1432

A DNA sequence (GBSx1518) was identified in *S.agalactiae* <SEQ ID 4397> which encodes the amino acid sequence <SEQ ID 4398>. This protein is predicted to be UDP-N-acetylglucosamine-1-carboxyvinyl transferase (murA). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.63   Transmembrane   25 -  41 (  24 -  42)

----- Final Results -----
               bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF86297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
           transferase [Listeria monocytogenes]
 Identities = 240/412 (58%), Positives = 303/412 (73%), Gaps = 2/412 (0%)

Query: 3   KIIINGGKQLTGEVAVSGAKNSVVALIPATILADDVVVLDGVPAISDVDSLVDIMETMGA 62
           K+II GGK+L G + V GAKNS VALIPA ILA+  VVL+G+P ISDV +L +I+E +G
Sbjct: 20  KLIIRGGKKLAGTLQVDGAKNSAVALIPAAILAESEVVLEGLPDISDVHTLYNILEELGG 79

Query: 63  KIKRYGETLEIDPCGVKDIPMPYGKINSLRASYYFYGSLLGRYGQATLGLPGGCDLGPRP 122
            ++   +T  IDP  +  +P+P G +  LRASYY  G++LGR+ +A +GLPGGC LGPRP
Sbjct: 80  TVRYDNKTAVIDPTDMISMPLPSGNVKKLRASYYLMGAMLGRFKKAVIGLPGGCYLGPRP 139

Query: 123 IDLHLKAFEAMGASVSYEGDSMRLATNGKPLQGANIYMDTVSVGATINTIIAAAKANGRT 182
           ID H+K FEA+GA V+ E  ++ L  +   L+GA IY+D VSVGATIN ++AA +A G+T
Sbjct: 140 IDQHIKGFEALGAKVTNEQGAIYLRAD--ELKGARIYLDVVSVGATINIMLAAVRAKGKT 197

Query: 183 VIENAAREPEIIDVATLLNNMGAHIRGAGTDVITIEGVKSLHGTRHQVIPDRIEAGTYIA 242
           VIENAA+EPEIIDVATLL NMGA I+GAGTD I I GV+ LHG  H +IPDRIEAGT++
Sbjct: 198 VIENAAKEPEIIDVATLLTNMGAIIKGAGTDTIRITGVEHLHGCHHTIIPDRIEAGTFMV 257

Query: 243 MAAAIGRGIKVTNVLYEHLESFIAKLDEMGVRMTVEEDSIFVEEQERLKAVSIKTSPYPG 302
           +AAA G+G+++ NV+  HLE  IAKL EMGV M +EED+IFV E E++K V IKT  YPG
Sbjct: 258 LAAASGKGVRIENVIPTHLEGIIAKLTEMGVPMDIEEDAIFVGEVEKIKKVDIKTYAYPG 317

Query: 303 FATDLQQPLTPLLLTAEGNGSLLDTIYEKRVNHVPELARMGANISTLGGKIVYSGPNQLS 362
           F TDLQQPLT LL  AEG+  + DTIY  R  H+ E+ RMG      G   V +GP QL
Sbjct: 318 FPTDLQQPLTALLTRAEGSSVITDTIYPSRFKHIAEIERMGGKFKLEGRSAVINGPVQLQ 377

Query: 363 GAPVKATDLRAGAALVIAGLMAEGRTEITNIEFILRGYSNIIEKLTSLGADI 414
           G+ V ATDLRAGAALVIA L+A+G TEI  +E I RGYS IIEKL+++GA+I
Sbjct: 378 GSKVTATDLRAGAALVIAALLADGETEIHGVEHIERGYSKIIEKLSAIGANI 429
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4399> which encodes the amino acid sequence <SEQ ID 4400>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.70   Transmembrane   25 -  41 (  23 -  45)

----- Final Results -----
               bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF86297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
           transferase [Listeria monocytogenes]
 Identities = 244/412 (59%), Positives = 302/412 (73%), Gaps = 2/412 (0%)
```

```
Query: 3   KIIINGGKALSGEVAVSGAKNSVVALIPAIILADDIVILDGVPAISDVDSLIEIMELMGA 62

                        K+II GGK L+G + V GAKNS VALIPA ILA+  V+L+G+P ISDV +L  I+E +G
Sbjct: 20  KLIIRGGKKLAGTLQVDGAKNSAVALIPAAILAESEVVLEGLPDISDVHTLYNILEELGG 79

Query: 63  TVNYHGDTLEIDPRGVQDIPMPYGKINSLRASYYFYGSLLGRFGQAVVGLPGGCDLGPRP 122
           TV Y   T  IDP  +  +P+P G +  LRASYY  G++LGRF +AV+GLPGGC LGPRP
Sbjct: 80  TVRYDNKTAVIDPTDMISMPLPSGNVKKLRASYYLMGAMLGRFKKAVIGLPGGCYLGPRP 139

Query: 123 IDLHLKAFEAMGVEVSYEGENMNLSTNGQKIHGAHIYMDTVSVGATINTMVAATKAQGKT 182
           ID H+K FEA+G +V+ E   + L  +  ++ GA IY+D VSVGATIN M+AA +A+GKT
Sbjct: 140 IDQHIKGFEALGAKVTNEQGAIYLRAD--ELKGARIYLDVVSVGATINIMLAAVRAKGKT 197

Query: 183 VIENAAREPEIIDVATLLNNMGAHIRGAGTDIITIQGVQKLHGTRHQVIPDRIEAGTYIA 242
           VIENAA+EPEIIDVATLL NMGA I+GAGTD I I GV+ LHG  H +IPDRIEAGT++
Sbjct: 198 VIENAAKEPEIIDVATLLTNMGAIIKGAGTDTIRITGVEHLHGCHHTIIPDRIEAGTFMV 257

Query: 243 LAAAIGKGVKITNVLYEHLESFIAKLEEMGVRMTVEEDAIFVEKQESLKAITIKTSPYPG 302
           LAAA GKGV+I NV+  HLE  IAKL EMGV M +EEDAIFV + E +K + IKT  YPG
Sbjct: 258 LAAASGKGVRIENVIPTHLEGIIAKLTEMGVPMDIEEDAIFVGEVEKIKKVDIKTYAYPG 317

Query: 303 FATDLQQPLTPLLLKADGRGTIIDTIYEKRINHVPELMRMGADISVIGGQIVYQGPSRLT 362
           F TDLQQPLT LL +A+G   I DTIY  R  H+ E+ RMG   + G   V  GP +L
Sbjct: 318 FPTDLQQPLTALLTRAEGSSVITDTIYPSRFKHIAEIERMGGKFKLEGRSAVINGPVQLQ 377

Query: 363 GAQVKATDLRAGAALVTAGLIAEGKTEITNIEFILRGYASIIAKLTALGADI 414
           G++V ATDLRAGAALV A L+A+G+TEI  +E I RGY+ II KL+A+GA+I
Sbjct: 378 GSKVTATDLRAGAALVIAALLADGETEIHGVEHIERGYSKIIEKLSAIGANI 429
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 344/419 (82%), Positives = 394/419 (93%)

Query: 1   MRKIIINGGKQLTGEVAVSGAKNSVVALIPATILADDVVVLDGVPAISDVDSLVDIMETM 60
           MRKIIINGGK L+GEVAVSGAKNSVVALIPA ILADD+V+LDGVPAISDVDSL++IME M
Sbjct: 1   MRKIIINGGKALSGEVAVSGAKNSVVALIPAIILADDIVILDGVPAISDVDSLIEIMELM 60

Query: 61  GAKIKRYGETLEIDPCGVKDIPMPYGKINSLRASYYFYGSLLGRYGQATLGLPGGCDLGP 120
           GA +  +G+TLEIDP GV+DIPMPYGKINSLRASYYFYGSLLGR+GQA +GLPGGCDLGP
Sbjct: 61  GATVNYHGDTLEIDPRGVQDIPMPYGKINSLRASYYFYGSLLGRFGQAVVGLPGGCDLGP 120

Query: 121 RPIDLHLKAFEAMGASVSYEGDSMRLATNGKPLQGANIYMDTVSVGATINTIIAAAKANG 180
           RPIDLHLKAFEAMG  VSYEG++M L+TNG+ + GA+IYMDTVSVGATINT++AA KA G
Sbjct: 121 RPIDLHLKAFEAMGVEVSYEGENMNLSTNGQKIHGAHIYMDTVSVGATINTMVAATKAQG 180

Query: 181 RTVIENAAREPEIIDVATLLNNMGAHIRGAGTDVITIEGVKSLHGTRHQVIPDRIEAGTY 240
           +TVIENAAREPEIIDVATLLNNMGAHIRGAGTD+ITI+GV+ LHGTRHQVIPDRIEAGTY
Sbjct: 181 KTVIENAAREPEIIDVATLLNNMGAHIRGAGTDIITIQGVQKLHGTRHQVIPDRIEAGTY 240

Query: 241 IAMAAAIGRGIKVTNVLYEHLESFIAKLDEMGVRMTVEEDSIFVEEQERLKAVSIKTSPY 300
           IA+AAAIG+G+K+TNVLYEHLESFIAKL+EMGVRMTVEED+IFVE+QE LKA++IKTSPY
Sbjct: 241 IALAAAIGKGVKITNVLYEHLESFIAKLEEMGVRMTVEEDAIFVEKQESLKAITIKTSPY 300

Query: 301 PGFATDLQQPLTPLLLTAEGNGSLLDTIYEKRVNHVPELARMGANISTLGGKIVYSGPNQ 360
           PGFATDLQQPLTPLLL A+G G+++DTIYEKR+NHVPEL RMGA+IS +GG+IVY GP++
Sbjct: 301 PGFATDLQQPLTPLLLKADGRGTIIDTIYEKRINHVPELMRMGADISVIGGQIVYQGPSR 360

Query: 361 LSGAPVKATDLRAGAALVIAGLMAEGRTEITNIEFILRGYSNIIEKLTSLGADIQLVEE 419
           L+GA VKATDLRAGAALV AGL+AEG+TEITNIEFILRGY++II KLT+LGADIQL+E+
Sbjct: 361 LTGAQVKATDLRAGAALVTAGLIAEGKTEITNIEFILRGYASIIAKLTALGADIQLIED 419
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1433

A DNA sequence (GBSx1519) was identified in *S.agalactiae* <SEQ ID 4401> which encodes the amino acid sequence <SEQ ID 4402>. This protein is predicted to be thiamine phosphate pyrophosphorylase (thiE). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0422(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25544 GB:AF109218 ThiE [Staphylococcus carnosus]
 Identities = 98/200 (49%), Positives = 140/200 (70%), Gaps = 1/200 (0%)

Query: 5   LKLYFVCGTVDCSR-KNILTVVEEALQAGITLFQFREKGFTALQGKEKIAMAKQLQILCK 63
           L +YF+CGT D    + I  V++EAL+ GITL+QFREKG  A  G++K+A+AK+LQ LCK
Sbjct: 7   LNVYFICGTQDIPEGRTIQEVLKEALEGGITLYQFREKGNGAKTGQDKVALAKELQALCK 66

Query: 64  QYQVPFIIDDDIDLVELIDADGLHIGQNDLPVDEARRRLPDKIIGLSVSTMDEYQKSQLS 123
            Y VPFI++DD+ L E IDADG+H+GQ+D  VD+   R   KIIGLS+  ++E   S L+
Sbjct: 67  SYNVPFIVNDDVALAEEIDADGIHVGQDDEAVDDFNNRFEGKIIGLSIGNLEELNASDLT 126

Query: 124 VVDYIGIGPFNPTQSKADAKPAVGNRTTKAVREINQDIPIVAIGGITSDFVHDIIESGAD 183
            VDYIG+GP   T SK DA   VG +  + +R+   D+PIVAIGGI+ D V ++ ++ AD
Sbjct: 127 YVDYIGVGPIFATPSKDDASEPVGPKMIETLRKEVGDLPIVAIGGISLDNVQEVAKTSAD 186

Query: 184 GIAVISAISKANHIVDATRQ 203
           G++VISAI+++ H+ +   +
Sbjct: 187 GVSVISAIARSPHVTETVHK 206
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1434

A DNA sequence (GBSx1520) was identified in *S.agalactiae* <SEQ ID 4403> which encodes the amino acid sequence <SEQ ID 4404>. This protein is predicted to be hydroxyethylthiazole kinase (b2104). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -4.94   Transmembrane  198 - 214 ( 194 - 217)

----- Final Results -----
               bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8805> which encodes amino acid sequence <SEQ ID 8806> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     -2.93
GvH: Signal Score (-7.5): 1.61
     Possible site: 39
>>> Seems to have no N-terminal signal sequence
```

```
ALOM program   count: 1 value:  -4.94 threshold:  0.0
   INTEGRAL    Likelihood = -4.94   Transmembrane  183 - 199 ( 179 - 202)
   PERIPHERAL  Likelihood =  2.49     151
 modified ALOM score:   1.49

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25543 GB:AF109218 ThiM [Staphylococcus carnosus]
 Identities = 114/253 (45%), Positives = 160/253 (63%), Gaps = 1/253 (0%)

Query: 18  LEQLKEVNPLTICITNNVVKNFTANGLLALGASPAMSECIEDLEDLLKVADALLINIGTL 77
           L+Q++  +PL IC TN+VVKNFTANGLL+LGASP MSE  ++ ED   VA ++LINIGTL
Sbjct: 5   LDQIRTEHPLVICYTNDVVKNFTANGLLSLGASPTMSEAPQEAEDFYPVAGSVLINIGTL 64

Query: 78  TKESWQLYQEAIKIANKNQVPVVLDPVAAGASRFRLEVSLDLLKNYSISLLTGNGSEIAA 137
           TK       E  KIAN+ + P+V DPVA GAS++R +     LK    +++ GN SEI A
Sbjct: 65  TKHHEHAMLENAKIANETETPLVFDPVAVGASKYRKDFCKYFLKKIKPTVIKGNASEILA 124

Query: 138 LIGEKQASKGADGGKVADLESIAVKANQVFDVPVVVTGETDAIAVRGEVRLLQNGSPLMP 197
           LI +    KG D     D+  IA KA + +   +++TGETD I    +V  L NGS  +
Sbjct: 125 LIDDTATMKGTDSADNLDVVDIAEKAYKEYQTAIILTGETDVIVQDNKVVKLSNGSHFLA 184

Query: 198 LVTGTGCLLGAVLAAFIGSSDRSDDLACLTEAMTVYNVAGEIAEKVAKGKGVGSFQVAFL 257
            +TG GCLLGAV+ AF+   +    +  L EA++VYN+A E AE+++  KG G+F   F+
Sbjct: 185 KITGAGCLLGAVVGAFL-FRNTHPSIETLIEAVSVYNIAAERAEQLSDSKGPGTFLTQFI 243

Query: 258 DALSQMKSEMIMD 270
           DAL ++ S+ + +
Sbjct: 244 DALYRIDSDAVAE 256
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8806 (GBS398) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 6; MW 31.8kDa).

The GBS398-His fusion product was purified (Figure 214, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 314), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1435

A DNA sequence (GBSx1521) was identified in *S.agalactiae* <SEQ ID 4405> which encodes the amino acid sequence <SEQ ID 4406>. This protein is predicted to be ThiD (thiD). Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25542 GB:AF109218 ThiD [Staphylococcus carnosus]
 Identities = 139/258 (53%), Positives = 186/258 (71%), Gaps = 4/258 (1%)

Query: 8   LTIAGTDPSGGAGIMADLKTFQARRTYGMAVVTSVVAQNTCGVRGVQHIETAIIDQQLAC 67
            LTIAGTDP+GGAG+MADLK+F A   YGMA +TS+VAQNT GV+ + +++   + +QL
Sbjct: 8   LTIAGTDPTGGAGVMADLKSFHACGVYGMAAITSIVAQNTKGVQHIHNLDITWLKEQLDS 67

Query: 68  VYDDIKPKAVKTGMLAERETISLVASYLKKYPQ-PYVLDPVMVATSGHRLIDSDAVEALK 126
            ++DD  P+A+KTGM+A +E + L+ SYL+KYP  PYV+DPVM+A SG  L+D     AL+
Sbjct: 68  IFDDELPQAIKTGMIATKEMMELIRSYLEKYPDIPYVIDPVMLAKSGDSLMDDAGKHALQ 127

Query: 127 EDLLPLATIITPNLPEAEVLVGYDLSDEVSIIKAGYDIQKQYSVRNVLIKGGHLD--GLA 184
            E LLPLA + TPNLPEAE +VG+ L  E +I KAG     +   + V+IKGGH++   +A
Sbjct: 128 EILLPLADVATPNLPEAEEIVGFKLDTEEAIKKAGDIFINEIGSKGVVIKGGHIEDKNIA 187

Query: 185 KDYLFLEKEGLITLSNQRINTIHTHGTGCTFAAVVAAELAKGQSILNAVSTAKSFITSAI 244
            KDYLF  K+GL    ++R +T HTHGTGCTF+AV+ AELAKG++I  AV  AK FI  +I
Sbjct: 188 KDYLF-TKDGLEVFESERYDTKHTHGTGCTFSAVITAELAKGKTIYEAVKKAKDFIALSI 246

Query: 245 ETAPELGLGNGPVNHTSY 262
            +  PE+G G GPVNH +Y
Sbjct: 247 KYTPEIGQGRGPVNHFAY 264
```

There is also homology to SEQ ID 4408.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1436

A DNA sequence (GBSx1522) was identified in *S.agalactiae* <SEQ ID 4409> which encodes the amino acid sequence <SEQ ID 4410>. This protein is predicted to be TenA (tenA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2242(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25541 GB:AF109218 TenA [Staphylococcus carnosus]
 Identities = 78/213 (36%), Positives = 127/213 (59%), Gaps = 6/213 (2%)

Query: 14  IQSIYQDPFIQGIIKGRLDHDVICHYLQADNIYLGKFADIYALCLAKSDNLRDKQFFLEQ 73
            I  IYQD FIQ ++KG +  + +  YL+AD  YL +FA+IYAL +    +L   +F ++Q
Sbjct: 15  IDEIYQDHFIQELLKGDIKKEALRQYLRADASYLREFANIYALLIPIMPDLESVRFLVDQ 74

Query: 74  IDFTLNRELADGEGPHQALAAYTNRSYQDIIEKGVWYPSADHYIKHMYFHFY-ENGIAGA 132
            I F +N E+      H+ +A Y   +Y +I++K VW PS DHYIKHMY++ Y     A A
Sbjct: 75  IQFIVNGEVE----AHEYMADYIGENYNEIVQKKVWPPSGDHYIKHMYYNVYAHENAAYA 130

Query: 133 LAAMSPCPWIYHQLAKKIIEENQFLNGNPFNNWITFYANDTVEELMENYFRMMDYYAQNL 192
            +AAM+PCP++Y  +AK+ +++      +    W  FY N  ++ L+E    +M+    N+
Sbjct: 131 IAAMAPCPYVYAMIAKRAMKDPNLNKSSILAKWFEFY-NTEMDPLIEVLDDLMNQLTANM 189

Query: 193 SKEKQADLVDAFVKSCQHERRFFQMAINQEKWE 225
            S+ ++ ++ + +++S  HE  FF MA   EKW+
Sbjct: 190 SETEKNEVRENYLQSTVHELNFFNMAYTSEKWQ 222
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1437

A DNA sequence (GBSx1523) was identified in *S.agalactiae* <SEQ ID 4411> which encodes the amino acid sequence <SEQ ID 4412>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.06   Transmembrane   43 -  59 (  36 -  63)
   INTEGRAL    Likelihood = -2.55   Transmembrane   92 - 108 (  92 - 112)
   INTEGRAL    Likelihood = -1.49   Transmembrane  135 - 151 ( 135 - 151)
   INTEGRAL    Likelihood = -1.06   Transmembrane   69 -  85 (  69 -  85)
   INTEGRAL    Likelihood = -0.22   Transmembrane  216 - 232 ( 216 - 232)

----- Final Results -----
              bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA91230 GB:Z56283 orf2 [Lactobacillus helveticus]
 Identities = 46/215 (21%), Positives = 96/215 (44%), Gaps = 3/215 (1%)

Query: 21  AITFLCLLIPTFSFSFTLRLRTSLLFLIIVVTLQCFVKVSLKTWAKVNLISFVMGLSLFL 80
           ++ F+   I +   S    L T+L+ +   +      ++ +K    +  + F+   ++F
Sbjct: 4   SLKFILAFIISLEISLKASLTTNLIVIAFALIYLLVTRIKIKELILLIAVPFIASFTIFA 63

Query: 81  GTYFWGKLPHQFVLASLVACRPLIFMNVGLLFHASHSNYDFIESLYQTFKVPSHFAYGIF 140
             +++   P  +   +L + R  ++         + +  DF  SL Q   +PS FAYG+
Sbjct: 64  TLFWFSPTPDAYYAWNL-STRVYVYTLTIACVTRNTTATDFARSLEQNLHLPSKFAYGVL 122

Query: 141 AVFNLLPLIKLQYQRNRLAFRLKNQVTWALSPRLILSVLLKTIYWVEQLELAMLSKGFEV 200
           A  N++P +K   ++ R +  ++       SP L    +L  +   + L   M S G+
Sbjct: 123 AAINIIPRMKTAVKQIRTSAMMRGMYLSFWSPVLYFKAILVALNSADNLAQGMESHGYVE 182

Query: 201 GKERTHASTYPVRFRDYSL-LGMSILLSIGM-IFK 233
           G++R      P+  +D+ +   + IL++I + IFK
Sbjct: 183 GQKRATIVAIPLTKKDWLIFFTLLILVNISLFIFK 217
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8807> and protein <SEQ ID 8808> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      4.50
GvH: Signal Score (-7.5): -0.2
     Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 5 value:  -7.06 threshold:  0.0
   INTEGRAL    Likelihood = -7.06   Transmembrane   43 -  59 (  36 -  63)
   INTEGRAL    Likelihood = -2.55   Transmembrane   92 - 108 (  92 - 112)
   INTEGRAL    Likelihood = -1.49   Transmembrane  135 - 151 ( 135 - 151)
   INTEGRAL    Likelihood = -1.06   Transmembrane   69 -  85 (  69 -  85)
   INTEGRAL    Likelihood = -0.22   Transmembrane  216 - 232 ( 216 - 232)
   PERIPHERAL  Likelihood =  2.65     170
 modified ALOM score:    1.91

*** Reasoning Step: 3
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1438

A DNA sequence (GBSx1524) was identified in *S.agalactiae* <SEQ ID 4413> which encodes the amino acid sequence <SEQ ID 4414>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3007(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA91229 GB:Z56283 orf1 [Lactobacillus helveticus]
 Identities = 123/424 (29%), Positives = 200/424 (47%), Gaps = 48/424 (11%)

Query: 17  LFDEVTFSLNPGERILISGYSGCGKSTLALLLSGL--KESGK--GQVLLNGSLIEPSDVG 72
            L +++  ++ PG  +LI G +GCGKSTL  +++GL  K +GK  G++ L+G
Sbjct: 12  LINQLNMNIAPGFNLLI-GPTGCGKSTLLKIIAGLYPKYAGKLTGKIDLHGQ-----KAA 65

Query: 73  FLFQNPDLQFCMDTVAHELYFILENLQIEPEQMQDRSEFVLAQVGLKGFQNRLIYTLSQG 132
             +FQN   QF M T   E+ F LENLQI+ +      +  +    +    ++ I TLS G
Sbjct: 66  MMFQNAAEQFTMTTPREEIIFALENLQIKAKDYDLHIKKAVEFTKIADLLDQKINTLSGG 125

Query: 133 EKQRLALATIFLKSPKLIILDEAFANLDQESASQLLQLVLNYQANNQSMLIVIDHLITYY 192
            ++Q +ALA +      + +LDE FA+ D  +   L++ + +     ++ +I+ DH++  Y
Sbjct: 126 QQQHVALAVLIAMDVDVFLLDEPFASCDPNTRHFLIEKLASLAETGRT-IILSDHVLDDY 184

Query: 193 QDIMDHYFWLEKRLTRVNFDYMLNRLNVFELEKKSHN--------TGDKLLSIKDFQVK- 243
            + I DH +  E +  +       N+L  F+  K+ H         TG  +  +   Q+K
Sbjct: 185 EKICDHLYQFEGKTVKELSANEKNKL--FKQNKQFHEQSYSFALPTGTPVFELNKTQIKQ 242

Query: 244 ----LSKNKFISYLDFDLASGERLCLDGPSGVGKSSLFMGLLGLYRTKGK--------KQ 291
                L +NK   Y       G+   + G +GVGK+SLF  +  +   KG          +
Sbjct: 243 NRLLLKQNKLKIY-------GKTTLITGSNGVGKTSLFKAMTKMIPYKGNFTYLDNEISK 295

Query: 292 FTHRKQIP-ISFLFQNPLDQFIFSTVYDEIFQVCKDSN------KARDILETINLWDKKQ 344
              +RK +  I+  FQ   DQF+  TV DEI    KD N      K  + LE + L
Sbjct: 296 IKYRKYLSQIAQFFQKASDQFLTVTVKDEIELSKKDRNNFFTDAKIDEWLEKLQLKQHLD 355

Query: 345 FSPFQLSQGQQRRLAIGSILASDSKLLLLDEPTYGQDAYHANMITTLLLSYCHKNHCGVI 404
               + LS GQQ++L I  +L +   +LL+DEP  G D   +++   L+     K     +
Sbjct: 356 QVVYSLSGGQQKKLQILLMLMTKHNVLLIDEPLSGLDHESVDLVLQLMQECQEKLQQTFL 415

Query: 405 FTSH 408
              SH
Sbjct: 416 IISH 419
 Identities = 44/185 (23%), Positives = 83/185 (44%), Gaps = 24/185 (12%)

Query: 28  GERILISGYSGCGKSTLALLLSGLKESGKGQVLLNGSLIEP------SDVGFLFQNPDLQ 81
            G+  LI+G +G GK++L   ++ +         L+  + +       S +   FQ    Q
Sbjct: 256 GKTTLITGSNGVGKTSLFKAMTKMIPYKGNFTYLDNEISKIKYRKYLSQIAQFFQKASDQ 315

Query: 82  FCMDTVAHELYFILENLQIEPEQMQDRSEFV--------LAQVGLKGFQNRLIYTLSQGE 133
            F   TV  E+              +DR+ F         L ++ LK   ++++Y+LS G+
```

```
Sbjct: 316 FLTVTVKDEIEL----------SKKDRNNFFTDAKIDEWLEKLQLKQHLDQVVYSLSGGQ 365

Query: 134 KQRLALATIFLKSPKLIILDEAFANLDQESASQLLQLVLNYQANNQSMLIVIDHLITYYQ 193
           +++L +  + +    ++++DE  + LD ES   +LQL+   Q   Q   ++I H I
Sbjct: 366 QKKLQILLMLMTKHNVLLIDEPLSGLDHESVDLVLQLMQECQEKLQQTFLIISHQIDALA 425

Query: 194 DIMDH 198
           D  D+
Sbjct: 426 DFCDY 430
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4415> which encodes the amino acid sequence <SEQ ID 4416>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3093(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 120/455 (26%), Positives = 203/455 (44%), Gaps = 47/455 (10%)

Query: 1   MLSVEKLACTHGDSHYLFDEV-TFSLNPGERILISGYSGCGKSTLALLLSGLKE---SGK 56
           M+S E+L  T+ D      ++ T  + G+ I++ G SG GKST   LL+G+     +GK
Sbjct: 21  MISAEQLVFTYHDQKNPACQISTCQIASGQFIVLCGPSGSGKSTFLKLLNGIIPDYYAGK 80

Query: 57  GQVLLNGSLIEPS---------DVGFLFQNPDLQFCMDTVAHELYFILENLQIEPEQMQD 107
            +  L+ +  +            V  +FQNP  QF    V HEL F  EN  ++ + +
Sbjct: 81  YEGRLDVADCQAGRDSVETFSRSVASVFQNPASQFFYREVQHELVFPCENQGLDAKVIMK 140

Query: 108 RSEFVLAQVGLKGFQNRLIYTLSQGEKQRLALATIFLKSPKLIILDEAFANLDQESASQL 167
           R   +          N+ ++ LS G+KQR+A+AT  ++   +++ DE  ANLD    + +
Sbjct: 141 RLWTLAEDFAFAELLNKDMFGLSGGQKQRVAIATAIMQGTNIMLFDEPTANLDSAGIAAV 200

Query: 168 LQLVLNYQANNQSMLIVIDHLITYYQDIMDHYFW-----LEKRLTRVNF---------DY 213
              +   +A  ++ +IV +H + Y  D+ D++F+     L  +LT  N          D
Sbjct: 201 KAYLTQLKAAGKT-IIVAEHRLHYLMDLADNFFYFKNGRLTDKLTTQNLLALTDEQRQDM 259

Query: 214 MLNRLNVFELE-------KKSHNTGDKLLSIKDFQVKLSKNKFISYLDFDLASGERLCLD 266
            L RL++ +L+       +  H   D  L I+   V+              A G    +
Sbjct: 260 GLRRLDLSDLKPVLAGKIESQHYRPDDSLCIEHLTVRAGSKILRCIEQLSFAVGSISGIT 319

Query: 267 GPSGVGKSSLFMGLLGLYRTKGKKQFTHRKQIPISFLFQNPLDQFIFSTVYDEIF--QVC 324
           G +G+GKS L   + G+     KK     + IP+S   +      +   V  ++F   V
Sbjct: 320 GSNGLGKSQLVYYIAGI--LDDKKATIKFQGIPLSAKQRLSKTSIVLQEVSLQLFAESVS 377

Query: 325 KDSN-------KARDILETINLWDKKQFSPFQLSQGQQRRLAIGSILASDSKLLLLDEPT 377
           K+ N       +  +++E ++L    +  P  LS G+Q+R+ I + L +D  +L+ DEP+
Sbjct: 378 KEVNLGHERHPRTTEVIERLSLTTLLERHPASLSGGEQQRVMIAASLLADKDILIFDEPS 437

Query: 378 YGQDAYHANMITTLLLSYCHKNHCGVIFTSHDPHL 412
            G D      +  LL+      H  VI  SHD  L
Sbjct: 438 SGLDLLQMKALANLLMQ-LKTQHKVVILISHDEEL 471
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1439

A DNA sequence (GBSx1525) was identified in *S.agalactiae* <SEQ ID 4417> which encodes the amino acid sequence <SEQ ID 4418>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.62   Transmembrane    8 -  24 (   1 -  30)
   INTEGRAL    Likelihood = -8.17   Transmembrane  145 - 161 ( 143 - 163)
   INTEGRAL    Likelihood = -6.32   Transmembrane   66 -  82 (  62 -  84)
   INTEGRAL    Likelihood = -3.77   Transmembrane  112 - 128 ( 111 - 132)
   INTEGRAL    Likelihood = -2.66   Transmembrane   43 -  59 (  43 -  59)

----- Final Results -----
              bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13180 GB:Z99110 ykoE [Bacillus subtilis]
 Identities = 68/177 (38%), Positives = 117/177 (65%), Gaps = 1/177 (0%)

Query: 5   LKDVLLIALLAVVLGVVYFGAGYISNAFVPFVGPIAHEVIYGIWFVAGPMALYILRKPGT 64
           +K++++++++++V  VVY    +  N      GPIA+E IYGIWF+   +A Y++RKPG
Sbjct: 6   VKEIVIMSVISIVFAVVYLLFTHFGNVLAGMFGPIAYEPIYGIWFIVSVIAAYMIRKPGA 65

Query: 65  AIVAELLAALIEVLIGSIYGPSVLVIGTLQGLGSELGFTLFRYHNYKLPAFILSAILTSI 124
           A+V+E++AAL+E L+G+  GP V+VIG +QGLG+E  F   R+  Y LP  +L+ + +S+
Sbjct: 66  ALVSEIIAALVECLLGNPSGPMVIVIGIVQGLGAEAVFLATRWKAYSLPVLMLAGMGSSV 125

Query: 125 FSFAWSFYANGLSAFSFSYNILMLIVRTVS-SIIFFLLTKNICDQLHRSGVLNAYGI 180
            SF +  + +G +A+S  Y ++ML++R +S +++  LL K +   L  +GVLN   +
Sbjct: 126 ASFIYDLFVSGYAAYSPGYLLIMLVIRLISGALLAGLLGKAVSGSLAYTGVLNGMAL 182
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1440

A DNA sequence (GBSx1526) was identified in *S.agalactiae* <SEQ ID 4419> which encodes the amino acid sequence <SEQ ID 4420>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.69   Transmembrane   65 -  81 (  53 -  95)
   INTEGRAL    Likelihood = -6.37   Transmembrane   34 -  50 (  31 -  54)
   INTEGRAL    Likelihood = -6.10   Transmembrane  176 - 192 ( 169 - 195)
   INTEGRAL    Likelihood = -3.66   Transmembrane  130 - 146 ( 130 - 151)
   INTEGRAL    Likelihood = -1.97   Transmembrane    3 -  19 (   3 -  19)
   INTEGRAL    Likelihood = -0.90   Transmembrane   88 - 104 (  88 - 104)

----- Final Results -----
              bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9757> which encodes amino acid sequence <SEQ ID 9758> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8809> and protein <SEQ ID 8810> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      -4.09
GvH: Signal Score (-7.5): -4.38
     Possible site: 47
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 6 value:  -6.69 threshold:  0.0
   INTEGRAL    Likelihood = -6.69   Transmembrane   65 -  81 (  53 -  95)
   INTEGRAL    Likelihood = -6.37   Transmembrane   34 -  50 (  31 -  54)
   INTEGRAL    Likelihood = -6.10   Transmembrane  176 - 192 ( 169 - 195)
   INTEGRAL    Likelihood = -3.66   Transmembrane  130 - 146 ( 130 - 151)
   INTEGRAL    Likelihood = -1.97   Transmembrane    3 -  19 (   3 -  19)
   INTEGRAL    Likelihood = -0.90   Transmembrane   88 - 104 (  88 - 104)
   PERIPHERAL  Likelihood =  5.30     158
 modified ALOM score:   1.84

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1441

A DNA sequence (GBSx1527) was identified in *S.agalactiae* <SEQ ID 4421> which encodes the amino acid sequence <SEQ ID 4422>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8811> and protein <SEQ ID 8812> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
McG: Discrim Score:      6.01
GvH: Signal Score (-7.5): 0.45
     Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  10.66 threshold:  0.0
   PERIPHERAL  Likelihood = 10.66      80
 modified ALOM score:  -2.63

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 4422 (GBS19) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 4; MW 24kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 6; MW 46.1kDa).

The GST-fusion protein was purified as shown in Figure 190, lane 10.

Example 1442

A DNA sequence (GBSx1528) was identified in *S.agalactiae* <SEQ ID 4423> which encodes the amino acid sequence <SEQ ID 4424>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8813> which encodes amino acid sequence <SEQ ID 8814> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length  of  UR:   23
     Peak Value of UR:   2.61
     Net Charge of CR: 3
McG: Discrim Score:      9.08
GvH: Signal Score (-7.5): -0.76
     Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 23
ALOM program   count: 0 value:   5.14 threshold:  0.0
   PERIPHERAL  Likelihood =  5.14     365
 modified ALOM score:  -1.53

*** Reasoning Step: 3

Rule gpo1

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA34476 GB:X16457 precursor polypeptide (AA -26 to 632)
           [Staphylococcus aureus]
 Identities = 93/372 (25%), Positives = 160/372 (43%), Gaps = 46/372 (12%)

Query: 9   MKKQFLKSAAILSLAVTAVSTSQPVGAIVGKDETKLRQQLGYIDSKKSGKKIDERWGEKI 68
           MKKQ +   A L++A +  +      AIV KD +K   +    +  K G  + + +  KI
Sbjct: 1   MKKQIISLGA-LAVASSLFTWDNKADAIVTKDYSK---ESRVNEKSKKGATVSDYYYWKI 56

Query: 69  YNYLSYELIEANEWINRSEFQEPEYRTILSEFKDKIDSIEYYLINLS----NIAKEDAHQ 124
            + L  +   A + +   ++ +P Y+        ++   + YL+        + K+
Sbjct: 57  IDSLEAQFTGAIDLLENYKYGDPIYKEAKDRLMTRVLGEDQYLLKKKIDEYELYKKWYKS 116

Query: 125 RNILQSLDKYEKSGIYNLDQGVYNYIYQEISSAKHKFSDGVDKIYRLDSTLFPFSVWYDK 184
            N   ++  + K  +YNL    YN I+  +  A ++F+  V +I   +  L  F
Sbjct: 117 SNKNTNMLTFHKYNLYNLTMNEYNDIFNSLKDAVYQFNKEVKEIEHKNVDLKQF------ 170

Query: 185 HLDNNDNYKDNKDFKEYIALLNEITRKARLGYQIVNNHKD-GEHKDEAEI-LDILIRDIT 242
```

```
                D   ++K  KE   L++EI       Y      KD GEH  E    LD+++ D
Sbjct: 171 -----DKDGEDKATKEVYDLVSEIDTLVVTYYA----DKDYGEHAKELRAKLDLILGDTD 221

Query: 243 FVSKDAPGYKYIPNKRIAAKIIEDLDGIINDFFKNTGKDKP-SLEKLKDTEFHKKYLNST 301
              K       I N+RI  ++I+DL+ II+DFF  T +++P S+ K   T+ + K  +
Sbjct: 222 NPHK-------ITNERIKKEMIDDLNSIIDDFFMETKQNRPNSITKYDPTKHNFKEKSEN 274

Query: 302 EPYSIETNLPSNYKELKEKQIKKLEYGYK-KSSKIY--TSAHYALYSEEIDAAKELLQKV 358
            +P     N     +E K K +K+ +  +K K+ K Y  T     +  EE    +  L KV
Sbjct: 275 KP-----NFDKLVEETK-KAVKEADESWKNKTVKKYEETVTKSPVVKEEKKVEEPQLPKV 328

Query: 359 KIAKDNYNEIKS 370
                 N  E+K+
Sbjct: 329 ----GNQQEVKT 336
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8814 (GBS119) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 2; MW 84.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 5; 2 bands).

The GBS119-GST fusion product was purified (Figure 109A; see also Figure 201, lane 6) and used to immunise mice (lane 1+2+3 product; 20μg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 109B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1443

A DNA sequence (GBSx1529) was identified in *S.agalactiae* <SEQ ID 4425> which encodes the amino acid sequence <SEQ ID 4426>. This protein is predicted to be s-adenosylmethionine synthetase (metK). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3609(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07019 GB:AP001518 S-adenosylmethionine synthetase [Bacillus halodurans]
 Identities = 266/390 (68%), Positives = 324/390 (82%), Gaps = 1/390 (0%)

Query: 4   RKLFTSESVSEGHPDKIADQISDAILDAILEQDPDAHVAAETAVYTGSVHVFGEISTTAY 63
           R+LFTSESV+EGHPDKI DQISD+ILD IL++DP+A VA ET+V TG V V GEI+T+ Y
Sbjct: 7   RRLFTSESVTEGHPDKICDQISDSILDEILKEDPNARVACETSVTTGLVLVAGEITTSTY 66

Query: 64  VDINRVVRNTIAEIGYDKAEYGFSAESVGVHPSLVEQSPDIAQGVNEALEVR-GSLEQDP 122
           VDI +VVR+TI  IGY +A+YGF +E+  V  S+ EQSPDIAQGVN+ALE R G +
Sbjct: 67  VDIPKVVRDTIRNIGYTRAKYGFDSETCAVLTSIDEQSPDIAQGVNQALEAREGQMTDAE 126

Query: 123 LDLIGAGDQGLMFGFAVDETPELMPLPISLAHQLVKKLTDLRKSGELTYLRPDAKSQVTV 182
           ++ IGAGDQGLMFG+A +ETPELMPLPISL+H+L ++L++ RK   L YLRPD K+QVTV
Sbjct: 127 IEAIGAGDQGLMFGYANNETPELMPLPISLSHKLARRLSEARKGEILPYLRPDGKTQVTV 186

Query: 183 EYDENDQPIRVDAVVISTQHDPNVTNDQLHKDVIEKVINEVIPSHYLDDQTKFFINPTGR 242
```

```
           EYDENDQ +R+D +VISTQH P VT +Q+  D+ + VI  V+P   +D++TK+FINPTGR
Sbjct: 187 EYDENDQSVRIDTIVISTQHHPEVTLEQIESDLKQHVIRSVVPEELIDEETKYFINPTGR 246

Query: 243 FVIGGPQGDSGLTGRKIIVDTYGGYSRHGGGAFSGKDATKVDRSASYAARYIAKNIVAAD 302
           FVIGGPQGD+GLTGRKIIVDTYGGY+RHGGGAFSGKD TKVDRS +YAARY+AKNIVAA
Sbjct: 247 FVIGGPQGDAGLTGRKIIVDTYGGYARHGGGAFSGKDPTKVDRSGAYAARYVAKNIVAAG 306

Query: 303 LAKKVEVQLAYAIGVAQPVSVRVDTFGTGVIAEADLEAAVRQIFDLRPAGIINMLDLKRP 362
           LA K EVQLAYAIGVA+PVS+ +DTFGTG ++EA L   VR+ FDLRPAGII MLDL+RP
Sbjct: 307 LADKCEVQLAYAIGVAKPVSISIDTFGTGQVSEARLVELVREHFDLRPAGIIKMLDLRRP 366

Query: 363 IYRQTAAYGHMGRTDIDLPWERVDKVQALK 392
           IY+QTAAYGH GRTD++LPWE+ DK + L+
Sbjct: 367 IYKQTAAYGHFGRTDVELPWEQTDKAEILR 396
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4427> which encodes the amino acid sequence <SEQ ID 4428>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3389(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 333/395 (84%), Positives = 361/395 (91%), Gaps = 1/395 (0%)

Query: 1   MSERKLFTSESVSEGHPDKIADQISDAILDAILEQDPDAHVAAETAVYTGSVHVFGEIST 60
           MSERKLFTSESVSEGHPDKIADQISDAILDAIL +DP+AHVAAET VYTGSVHVFGEIST
Sbjct: 1   MSERKLFTSESVSEGHPDKIADQISDAILDAILAEDPEAHVAAETCVYTGSVHVFGEIST 60

Query: 61  TAYVDINRVVRNTIAEIGYDKAEYGFSAESVGVHPSLVEQSPDIAQGVNEALEVRGSLEQ 120
           TAY+DINRVVR+TIAEIGY +AEYGFSAESVGVHPSLVEQS DIAQGVNEA E R   +
Sbjct: 61  TAYIDINRVVRDTIAEIGYTEAEYGFSAESVGVHPSLVEQSGDIAQGVNEAFESREG-DT 119

Query: 121 DPLDLIGAGDQGLMFGFAVDETPELMPLPISLAHQLVKKLTDLRKSGELTYLRPDAKSQV 180
           D L  IGAGDQGLMFGFA++ETPELMPLPISL+HQLV++L +LRKSGE++YLRPDAKSQV
Sbjct: 120 DDLSHIGAGDQGLMFGFAINETPELMPLPISLSHQLVRRLAELRKSGEISYLRPDAKSQV 179

Query: 181 TVEYDENDQPIRVDAVVISTQHDPNVTNDQLHKDVIEKVINEVIPSHYLDDQTKFFINPT 240
           TVEYDE+D+P+RVD VVISTQHDP  TNDQ+ +DVIEKVI  VIP+ YLDD TKFFINPT
Sbjct: 180 TVEYDEHDKPVRVDTVVISTQHDPEATNDQIRQDVIEKVIKAVIPADYLDDDTKFFINPT 239

Query: 241 GRFVIGGPQGDSGLTGRKIIVDTYGGYSRHGGGAFSGKDATKVDRSASYAARYIAKNIVA 300
           GRFVIGGPQGDSGLTGRKIIVDTYGGYSRHGGGAFSGKDATKVDRSASYAARYIAKN+VA
Sbjct: 240 GRFVIGGPQGDSGLTGRKIIVDTYGGYSRHGGGAFSGKDATKVDRSASYAARYIAKNLVA 299

Query: 301 ADLAKKVEVQLAYAIGVAQPVSVRVDTFGTGVIAEADLEAAVRQIFDLRPAGIINMLDLK 360
           A L  K EVQLAYAIGVAQPVSVRVDTFGT  + EA LEAAVRQ+FDLRPAGII MLDLK
Sbjct: 300 AGLVTKAEVQLAYAIGVAQPVSVRVDTFGTSTVPEAVLEAAVRQVFDLRPAGIIQMLDLK 359

Query: 361 RPIYRQTAAYGHMGRTDIDLPWERVDKVQALKDFI 395
           RPIY+QTAAYGHMGRTDIDLPWER++KV AL + +
Sbjct: 360 RPIYKQTAAYGHMGRTDIDLPWERLNKVDALVEAV 394
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1444

A DNA sequence (GBSx1530) was identified in *S.agalactiae* <SEQ ID 4429> which encodes the amino acid sequence <SEQ ID 4430>. This protein is predicted to be a transcriptional repressor of the biotin operon. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  188 - 204 ( 188 - 204)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9755> which encodes amino acid sequence <SEQ ID 9756> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
           operon [Bacillus halodurans]
 Identities = 102/315 (32%), Positives = 169/315 (53%), Gaps = 18/315 (5%)

Query: 10  ILSKNNNFISGETMANQLNISRTAIWKGIKTLEELGLEIESVTNKGYRLVSG-DILLPEQ 68
           +L+  ++F+SGE ++  +  SRTA+WK I+ L + G E+E+V  KGYR+V   D + P
Sbjct: 9   LLTAGDDFVSGEKISQAIGCSRTAVWKHIEELRKSGYEVEAVQRKGYRIVKRPDQIKPHD 68

Query: 69  LE-----QEIGIKVSLNNNSASTQLDAKMGIESKLKTPHLFLAPNQKKAKGRFDRPFFTS 123
           ++     +  G +++   ++ASTQ  A    +   K  H+ LA  Q   KGR  R +++
Sbjct: 69  IQVVLETERFGREITYLESTASTQTVALKLAQEGAKEGHIVLANEQTSGKGRMGRGWYSP 128

Query: 124 NQGGIYMSLLLQPNVPIEDIKPYTVMVASSAVKAISRLTGITPEIKWVNDIYLDNKKIAG 183
               I MS++ +P +P +     T++ A + V+AI   TG+  +IKW ND+ +D KKI G
Sbjct: 129 PGSSISMSIIFRPQLPPQKAPQLTLLTAVAIVRAIKETTGLDSDIKWPNDLLIDGKKIVG 188

Query: 184 ILTEAIASVESGLVTNVIIGLGINFYIKE--FPRALTKRAGSLFTEQ-PTITRNQLITEI 240
           ILTE  A  +S  V +VI G+GIN   +E  F   + K A SL  ++   I R  LI  I
Sbjct: 189 ILTEMQADQDS--VHSVIQGIGINVNHQEEAFAEEIRKIATSLAIKKGEPIQRAPLIAAI 246

Query: 241 W---NLFFNIPLEDHLK----VYREKSLVLDRTVSFMDGQTMYSGKAIDITDKGYLVVEL 293
                LF+++ L+        ++   ++ + + +       +  G A  ITD G L++E
Sbjct: 247 LKNIELFYDLYLQHGFSRIKPLWEAHAISIGKRIRARMLNDVKFGVAKGITDDGVLLLED 306

Query: 294 DDGQLKTLRSGEISL 308
           DDG+L ++ S +I +
Sbjct: 307 DDGKLHSIYSADIEI 321
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4431> which encodes the amino acid sequence <SEQ ID 4432>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  194 - 210 ( 194 - 211)

----- Final Results -----
               bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
           operon [Bacillus halodurans]
 Identities = 98/315 (31%), Positives = 165/315 (52%), Gaps = 18/315 (5%)
```

```
Query: 10  LLSQTDDFVSGEYLADQLSISRTSVWKSIKSLENQGIQIDSLKHKGYRMVQG-DILLPKT 68
           LL+  DDFVSGE ++  +  SRT+VWK I+ L   G ++++++ KGYR+V+  D + P
Sbjct: 9   LLTAGDDFVSGEKISQAIGCSRTAVWKHIEELRKSGYEVEAVQRKGYRIVKRPDQIKPHD 68

Query: 69  I-----SQGLGMPVTYTPHSQSTQLDAKQGIEAHNSAPRLYLAPSQEAAKGRLDRQFFSA 123
           I     ++  G  +TY   + STQ  A +  +       + LA  Q + KGR+ R ++S
Sbjct: 69  IQVVLETERFGREITYLESTASTQTVALKLAQEGAKEGHIVLANEQTSGKGRMGRGWYSP 128

Query: 124 STGGIYMSMYLKPNVPYADMPPYTMMVASSIVKAISRLTGIDTEIKWVNDIYLGNHKVAG 183
               I MS+  +P +P    P  T++ A +IV+AI   TG+D++IKW ND+ +   K+ G
Sbjct: 129 PGSSISMSIIFRPQLPPQKAPQLTLLTAVAIVRAIKETTGLDSDIKWPNDLLIDGKKIVG 188

Query: 184 ILTEAITSVETGLITDVIIGVGLNFFVTD--FPEAIAQKAGSLFTEK-PTITRNDLIIDI 240
           ILTE     +   +  VI G+G+N    +  F E I + A SL  +K   I R  LI  I
Sbjct: 189 ILTE--MQADQDSVHSVIQGIGINVNHQEEAFAEEIRKIATSLAIKKGEPIQRAPLIAAI 246

Query: 241 WK-------LFLSIPVKDHVKVYKEKSLVLNKQVTFIENSQEKRAIAIDLTDQGHLIVQF 293
            K       L+L         +++  ++ + K++     +  K  +A  +TD G L+++
Sbjct: 247 LKNIELFYDLYLQHGFSRIKPLWEAHAISIGKRIRARMLNDVKFGVAKGITDDGVLLLED 306

Query: 294 ENGDLQTLRSGEISL 308
           ++G L ++ S +I +
Sbjct: 307 DDGKLHSIYSADIEI 321
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 191/311 (61%), Positives = 257/311 (82%)

Query: 1   MKTYEKIYQILSKNNNFISGETMANQLNISRTAIWKGIKTLEELGLEIESVTNKGYRLVS 60
           MKT EKIYQ+LS+ ++F+SGE +A+QL+ISRT++WK IK+LE  G++I+S+ +KGYR+V
Sbjct: 1   MKTSEKIYQLLSQTDDFVSGEYLADQLSISRTSVWKSIKSLENQGIQIDSLKHKGYRMVQ 60

Query: 61  GDILLPEQLEQEIGIKVSLNNNSASTQLDAKMGIESKLKTPHLFLAPNQKKAKGRFDRPF 120
           GDILLP+ + Q +G+ V+   +S STQLDAK GIE+    P L+LAP+Q+ AKGR DR F
Sbjct: 61  GDILLPKTISQGLGMPVTYTPHSQSTQLDAKQGIEAHNSAPRLYLAPSQEAAKGRLDRQF 120

Query: 121 FTSNQGGIYMSLLLQPNVPIEDIKPYTVMVASSAVKAISRLTGITPEIKWVNDIYLDNKK 180
           F+++ GGIYMS+ L+PNVP  D+ PYT+MVASS VKAISRLTGI  EIKWVNDIYL N K
Sbjct: 121 FSASTGGIYMSMYLKPNVPYADMPPYTMMVASSIVKAISRLTGIDTEIKWVNDIYLGNHK 180

Query: 181 IAGILTEAIASVESGLVTNVIIGLGINFYIKEFPRALTKRAGSLFTEQPTITRNQLITEI 240
           +AGILTEAI SVE+GL+T+VIIG+G+NF++ +FP A+ ++AGSLFTE+PTITRN LI +I
Sbjct: 181 VAGILTEAITSVETGLITDVIIGVGLNFFVTDFPEAIAQKAGSLFTEKPTITRNDLIIDI 240

Query: 241 WNLFFNIPLEDHLKVYREKSLVLDRTVSFMDGQTMYSGKAIDITDKGYLVVELDDGQLKT 300
           W LF +IP++DH+KVY+EKSLVL++ V+F++        AID+TD+G+L+V+ ++G L+T
Sbjct: 241 WKLFLSIPVKDHVKVYKEKSLVLNKQVTFIENSQEKRAIAIDLTDQGHLIVQFENGDLQT 300

Query: 301 LRSGEISLSSW 311
           LRSGEISLSSW
Sbjct: 301 LRSGEISLSSW 311
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1445

A DNA sequence (GBSx1531) was identified in *S.agalactiae* <SEQ ID 4433> which encodes the amino acid sequence <SEQ ID 4434>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.76    Transmembrane    3 -  19 (   3 -  20)

----- Final Results -----
```

```
                 bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1446

A DNA sequence (GBSx1532) was identified in *S.agalactiae* <SEQ ID 4435> which encodes the amino acid sequence <SEQ ID 4436>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -2.28   Transmembrane   24 -  40 (  24 -  40)

----- Final Results -----
                 bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4437> which encodes the amino acid sequence <SEQ ID 4438>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -1.91   Transmembrane   58 -  74 (  58 -  75)

----- Final Results -----
                 bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 37/67 (55%), Positives = 54/67 (80%), Gaps = 3/67 (4%)

Query: 1   MTKRQFIFMALLCSFETYFFNQSVMDGSWIFAIFWGVLLLRDLQKVYAISKFTKELIK-- 58
           MT RQF+FMA +C+FETYFFN  ++ G+++FA+FWG+LL RDL++V+ I++ TK ++K
Sbjct: 36  MTIRQFLFMAFVCAFETYFFNDLLLSGNYLFALFWGLLLFRDLRRVHTINQLTKTILKTA 95

Query: 59  -STKKKD 64
            S KKKD
Sbjct: 96  NSPKKKD 102
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1447

A DNA sequence (GBSx1533) was identified in *S.agalactiae* <SEQ ID 4439> which encodes the amino acid sequence <SEQ ID 4440>. This protein is predicted to be DNA polymerase III, gamma subunit (dnaZX). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1567(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4441> which encodes the amino acid sequence <SEQ ID 4442>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane  232 - 248 ( 232 - 249)

----- Final Results -----
               bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 408/558 (73%), Positives = 473/558 (84%), Gaps = 6/558 (1%)

Query: 1   MYQALYRKYRSQTFDEMVGQSVISTTLKQAVSSKKISHAYLFSGPRGTGKTSAAKIFAKA 60
           MYQALYRKYRSQTFDEMVGQSVISTTLKQAV S KISHAYLFSGPRGTGKTSAAKIFAKA
Sbjct: 1   MYQALYRKYRSQTFDEMVGQSVISTTLKQAVESGKISHAYLFSGPRGTGKTSAAKIFAKA 60

Query: 61  MNCPNQINGEPCNHCDICRDITNGSLEDVIEIDAASNNGVDEIRDIRDKSTYAPSRATYK 120
           MNCPNQ++GEPCN CDICRDITNGSLEDVIEIDAASNNGVDEIRDIRDKSTYAPSRATYK
Sbjct: 61  MNCPNQVDGEPCNQCDICRDITNGSLEDVIEIDAASNNGVDEIRDIRDKSTYAPSRATYK 120

Query: 121 VYIIDEVHMLSTGAFNALLKTLEEPTENVVFILATTELHKIPATILSRVQRFEFKAIKLL 180
           VYIIDEVHMLSTGAFNALLKTLEEPTENVVFILATTELHKIPATILSRVQRFEFKAIK
Sbjct: 121 VYIIDEVHMLSTGAFNALLKTLEEPTENVVFILATTELHKIPATILSRVQRFEFKAIKQK 180

Query: 181 AIRDHLAQILDKEAISYDLDALTLVARRAEGGMRDALSILDQALSLAKDNHISLDVAEEI 240
           AIR+HLA +LDKE I+Y++DAL L+ARRAEGGMRDALSILDQALSL+ DN +++  +AEEI
Sbjct: 181 AIREHLAWVLDKEGIAYEVDALNLIARRAEGGMRDALSILDQALSLSPDNQVAIAIAEEI 240

Query: 241 TGSISLSAIDDYVSNILAHDTTEALAKLEVIFDSGKSMSRFATDLLMYLRDLLVVQAGGE 300
           TGSIS+ A+ DYV  +     T+ALA LE I+DSGKSMSRFATDLL YLRDLLVV+AGG+
Sbjct: 241 TGSISILALGDYVRYVSQEQATQALAALETIYDSGKSMSRFATDLLTYLRDLLVVKAGGD 300

Query: 301 DSHSSDTFIANLNVKQDILFEMIDKVTSVLPEIKNGSHPKVYAEMMTIQLSEMVEKNSS- 359
           +   S  F  NL++  D +F+MI  VTS LPEIK G+HP++YAEMMTIQL++  +  S
Sbjct: 301 NQRQSAVFDTNLSLSIDRIFQMITVVTSHLPEIKKGTHPRIYAEMMTIQLAQKEQILSQV 360

Query: 360 NIPADVTAELDSLRRELKSLKNEMSQL-SRADQSSSTQKVKVNNKTFTFKVDRTKILTIM 418
           N+  ++ +E+++L+ EL  LK ++SQL SR D  + + K K   KT +++VDR  IL IM
Sbjct: 361 NLSGELISEIETLKNELAQLKQQLSQLQSRPDSLARSDKTK--PKTTSYRVDRVTILKIM 418

Query: 419 EETVVDSQRSREYLEALKSAWNEILDNITAQDRALLMGSEPVLANSENAILAFDAAFNAE 478
           EETV +SQ+SR+YL+ALK+AWNEILDNI+AQDRALLMGSEPVLANSENAILAF+AAFNAE
Sbjct: 419 EETVRNSQQSRQYLDALKNAWNEILDNISAQDRALLMGSEPVLANSENAILAFEAAFNAE 478

Query: 479 QAMKRTDLNDIFGNIMSKAAGFSPNILAVPRNDFNQIRSDFAKKMKAQK--TETEPEVNH 536
           Q M R +LND+FGNIMSKAAGFSPNILAVPR DF  IR +FA++MK+QK   + E EV
Sbjct: 479 QVMSRNNLNDMFGNIMSKAAGFSPNILAVPRTDFQHIRKEFAQQMKSQKDSVQEEQEVAL 538

Query: 537 QIPEDFSYLAERIAIVED 554
            IPE F +L ++I  ++D
Sbjct: 539 DIPEGFDFLLDKINTIDD 556
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1448

A DNA sequence (GBSx1534) was identified in *S.agalactiae* <SEQ ID 4443> which encodes the amino acid sequence <SEQ ID 4444>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence (or aa 1-19)

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06927 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 67/143 (46%), Positives = 96/143 (66%)

Query: 8   ENYQLLLLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLFDGEELILGPFQGGV 67
           E Y L+  Q  AL   E++A+ANL+NASA+L   L    + GFYL    EL+LGPFQG
Sbjct: 13  EKYSLVTKQLAALLEGESDAIANLANASALLYHFLEEVNWVGFYLIKEGELVLGPFQGLP 72

Query: 68  SCVHITLGKGVCGESAQTAKTLIVDDVTKHANYISCDSKAMSEIVVPMFKNGKLLGVLDL 127
           +CV I +G+GVCG +A+  +T+ V+DV +   +I+CD+ + SEIV+P+F+NG L GVLD+
Sbjct: 73  ACVRIPIGRGVCGTAAKEEQTVRVEDVHQFPGHIACDAASRSEIVIPLFQNGVLYGVLDI 132

Query: 128 DSSLVADYDEIDQEYLEKFVGIL 150
           DS  +  + E +Q  LE FV +L
Sbjct: 133 DSPSLNRFSEEEQALLESFVDVL 155
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4445> which encodes the amino acid sequence <SEQ ID 4446>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1753(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 122/164 (74%), Positives = 144/164 (87%)

Query: 1   MNKSKKIENYQLLLLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLFDGEELIL 60
           MNKSKKIE YQL++ QA+ LF++E+NALANLSNASA+LN  LPNSVFTGFYLFDG+ELIL
Sbjct: 1   MNKSKKIEQYQLMIAQAKELFANESNALANLSNASALLNMTLPNSVFTGFYLFDGQELIL 60

Query: 61  GPFQGGVSCVHITLGKGVCGESAQTAKTLIVDDVTKHANYISCDSKAMSEIVVPMFKNGK 120
           GPFQG VSCVHI LGKGVCGESAQ+ +T+I++DV +HANYISCD+ AMSEIVVPM K G
Sbjct: 61  GPFQGRVSCVHIKLGKGVCGESAQSRRTIIINDVKQHANYISCDAAAMSEIVVPMVKEGH 120

Query: 121 LLGVLDLDSSLVADYDEIDQEYLEKFVGILVEHTIWNLDMFGVE 164
           L+GVLDLDSSLVADYDE+DQEYLE FV + +E T +  +MFGV+
Sbjct: 121 LIGVLDLDSSLVADYDEVDQEYLEAFVDLFLEKTTFTFNMFGVK 164
```

SEQ ID 4444 (GBS282) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 9; MW 19.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 6; MW 44.8kDa) and in Figure 63 (lane 7; MW 47kDa).

The GBS282-GST fusion product was purified (Figure 211, lane 4; see also Figure 225, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 269), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1449

A DNA sequence (GBSx1535) was identified in *S.agalactiae* <SEQ ID 4447> which encodes the amino acid sequence <SEQ ID 4448>. This protein is predicted to be uridine kinase (udk). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14675 GB:Z99117 uridine kinase [Bacillus subtilis]
 Identities = 133/207 (64%), Positives = 167/207 (80%)

Query: 1   MRKKPIIIGVTGGSGGGKTSVSRAILSNFPDQKITMIEHDSYYKDQSHLTFEERVKTNYD 60
           M K P++IG+ GGSG GKTSV+R+I   F    I MI+ D YYKDQSHL FEER+ TNYD
Sbjct: 1   MGKNPVVIGIAGGSGSGKTSVTRSIYEQFKGHSILMIQQDLYYKDQSHLPFEERLNTNYD 60

Query: 61  HPLAFDTNLMIEQLNELIEGRPVDIPVYDYTKHTRSDRTIRQEPQDVIIVEGILVLEDQR 120
           HPLAFD + +IE + +L+  RP++ P+YDY  HTRS+ T+  EP+DVII+EGILVLED+R
Sbjct: 61  HPLAFDNDYLIEHIQDLLNYRPIEKPIYDYKLHTRSEETVHVEPKDVIILEGILVLEDKR 120

Query: 121 LRDLMDIKLFVDTDDDIRIIRRIKRDMEERDRSLDSIIEQYTEVVKPMYHQFIEPTKRYA 180
           LRDLMDIKL+VDTD D+RIIRRI RD+ ER RS+DS+IEQY  VV+PM++QF+EPTKRYA
Sbjct: 121 LRDLMDIKLYVDTDADLRIIRRIMRDINERGRSIDSVIEQYVSVVRPMHNQFVEPTKRYA 180

Query: 181 DIVIPEGVSNIVAIDLINTKVASILNE 207
           DI+IPEG  N VAIDL+ TK+ +IL +
Sbjct: 181 DIIIPEGGQNHVAIDLMVTKIQTILEQ 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4449> which encodes the amino acid sequence <SEQ ID 4450>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9151> which encodes the amino acid sequence <SEQ ID 9152>. Analysis of this protein sequence reveals the following:

```
      Possible site: 35
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 173/207 (83%), Positives = 193/207 (92%)

Query: 1   MRKKPIIIGVTGGSGGGKTSVSRAILSNFPDQKITMIEHDSYYKDQSHLTFEERVKTNYD 60
           M KKPIIIGVTGGSGGGKTSVSRAIL +FP+ +I MI+HDSYYKDQSH++FEERVKTNYD
Sbjct: 5   MLKKPIIIGVTGGSGGGKTSVSRAILDSFPNARIAMIQHDSYYKDQSHMSFEERVKTNYD 64

Query: 61  HPLAFDTNLMIEQLNELIEGRPVDIPVYDYTKHTRSDRTIRQEPQDVIIVEGILVLEDQR 120
           HPLAFDT+ MI+QL EL+ GRPVDIP+YDY KHTRS+ T RQ+PQDVIIVEGILVLED+R
Sbjct: 65  HPLAFDTDFMIQQLKELLAGRPVDIPIYDYKKHTRSNTTFRQDPQDVIIVEGILVLEDER 124

Query: 121 LRDLMDIKLFVDTDDDIRIIRRIKRDMEERDRSLDSIIEQYTEVVKPMYHQFIEPTKRYA 180
           LRDLMDIKLFVDTDDDIRIIRRIKRDM ER RSL+SII+QYT VVKPMYHQFIEP+KRYA
Sbjct: 125 LRDLMDIKLFVDTDDDIRIIRRIKRDMMERGRSLESIIDQYTSVVKPMYHQFIEPSKRYA 184

Query: 181 DIVIPEGVSNIVAIDLINTKVASILNE 207
           DIVIPEGVSN+VAID+IN+K+ASIL E
Sbjct: 185 DIVIPEGVSNVVAIDVINSKIASILGE 211
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1450

A DNA sequence (GBSx1536) was identified in *S.agalactiae* <SEQ ID 4451> which encodes the amino acid sequence <SEQ ID 4452>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5083(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12572 GB:Z99108 similar to RNA helicase [Bacillus subtilis]
 Identities = 140/343 (40%), Positives = 202/343 (58%), Gaps = 9/343 (2%)

Query: 10  QDKLTQRQFDDLTDIQNKLFQPITDGDNILGISPTGTGKTLAYLFPTLLKLQPK-KSQQL 68
           Q+      F   T +Q +  Q I DG +++  SPTGTGKTLAY  P L +++P+ K  Q
Sbjct: 16  QENWNASGFQKPTPVQEQAAQLIMDGKDVIAESPTGTGKTLAYALPVLERIKPEQKHPQA 75

Query: 69  LILAPNSELAGQIFDVTKEWAEPLGLTAQLFLSGSSQKRQIERLKKGPEILIGTAGRVFE 128
           +ILAP+ EL  QIF V ++W     L A   + G++ K+Q+E+LKK P I++GT GRVFE
Sbjct: 76  VILAPSRELVMQIFQVIQDWKAGSELRAASLIGGANVKKQVEKLKKHPHIIVGTPGRVFE 135

Query: 129 LVKLKKIKMMNINTIVLDEFDELLGDSQYHFVDNIINRVPRDQQMIYISATNKLDNS--- 185
           L+K KK+KM  + TIVLDE D+L+       +  II    RD+Q++  SAT K +
Sbjct: 136 LIKAKKLKMHEVKTIVLDETDQLVLPEHRETMKQIIKTTLRDRQLLCFSATLKKETEDVL 195

Query: 186 -KLADNTITIDLSNQKLDT--IKHYYITVDKRERTDLLRKFSNIPDFRGLVFFNSLSDLG 242
            +LA     + +   K +   +KH Y+  D+R++  LL+K S +   + LVF   + +L
Sbjct: 196 RELAQEPEVLKVQRSKAEAGKVKHQYLICDQRDKVKLLQKLSRLEGMQALVFVRDIGNLS 255

Query: 243 ACEERLQFNRASAVSLASDINIKFRKVILEKFKNHDISLLLGTDLVARGIDIDNLEYVIN 302
              E+L ++      L S+     R  I+  F++ +  LLL TD+ ARG+DI+NL YVI+
Sbjct: 256 VYAEKLAYHHVELGVLHSEAKKMERAKIIATFEDGEFPLLLATDIAARGLDIENLPYVIH 315

Query: 303 FDIARDKETYTHRSGRTGRMGKEGCVITFVTHKEELKQLKKYA 345
            DI  D++ Y HRSGRTGR GKEG V++ VT  EE K LKK A
Sbjct: 316 ADIP-DEDGYVHRSGRTGRAGKEGNVLSLVTKLEESK-LKKMA 356
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4453> which encodes the amino acid sequence <SEQ ID 4454>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3847(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 273/358 (76%), Positives = 312/358 (86%)

Query: 1   MITKFPDQWQDKLTQRQFDDLTDIQNKLFQPITDGDNILGISPTGTGKTLAYLFPTLLKL 60
           MITKFP QWQ+KL Q  F  LT IQ + FQPI DG N LGISPTGTGKTLAY+FP LL L
Sbjct: 12  MITKFPPQWQEKLDQVAFTHLTPIQEQAFQPIVDGKNFLGISPTGTGKTLAYVFPNLLAL 71

Query: 61  QPKKSQQLLILAPNSELAGQIFDVTKEWAEPLGLTAQLFLSGSSQKRQIERLKKGPEILI 120
            PKKSQQLLILAPN+ELAGQIF+VTK+WA+PLGLTAQLF+SG+SQKRQIERLKKGPEILI
Sbjct: 72  TPKKSQQLLILAPNTELAGQIFEVTKDWAQPLGLTAQLFISGTSQKRQIERLKKGPEILI 131

Query: 121 GTAGRVFELVKLKKIKMMNINTIVLDEFDELLGDSQYHFVDNIINRVPRDQQMIYISATN 180
           GT GR+FEL+KLKKIKMM++NTIVLDE+DELLGDSQY FV  I + VPRD QM+Y+SATN
Sbjct: 132 GTPGRIFELIKLKKIKMMSVNTIVLDEYDELLGDSQYDFVQKISHYVPRDHQMVYMSATN 191

Query: 181 KLDNSKLADNTITIDLSNQKLDTIKHYYITVDKRERTDLLRKFSNIPDFRGLVFFNSLSD 240
           K+D + LA NT  IDLS Q  D I+H+Y+ VDKRERTDLLRKF+NIP FR LVFFNSLSD
Sbjct: 192 KVDQTSLAPNTFCIDLSEQTNDAIQHFYLMVDKRERTDLLRKFTNIPHFRALVFFNSLSD 251

Query: 241 LGACEERLQFNRASAVSLASDINIKFRKVILEKFKNHDISLLLGTDLVARGIDIDNLEYV 300
           LGA EERLQ+N A+AVSLASDIN+KFRK ILEKFK+H +SLLL TDLVARGIDIDNL+YV
Sbjct: 252 LGATEERLQYNGAAAVSLASDINVKFRKTILEKFKSHQLSLLLATDLVARGIDIDNLDYV 311

Query: 301 INFDIARDKETYTHRSGRTGRMGKEGCVITFVTHKEELKQLKKYATVTELVLHNQKLH 358
           I+FD+ARDKE YTHR+GRTGRMGK G VITFV+H E+LK+LKK+A V+E+ L NQ+LH
Sbjct: 312 IHFDVARDKENYTHRAGRTGRMGKSGIVITFVSHPEDLKKLKKFAKVSEISLKNQQLH 369
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1451

A DNA sequence (GBSx1537) was identified in *S.agalactiae* <SEQ ID 4455> which encodes the amino acid sequence <SEQ ID 4456>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.38    Transmembrane   15 -  31 (  13 -  31)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1452

A DNA sequence (GBSx1538) was identified in *S.agalactiae* <SEQ ID 4457> which encodes the amino acid sequence <SEQ ID 4458>. This protein is predicted to be peptidoglycan GlcNAc deacetylase. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.92   Transmembrane    4 -  20 (   1 -  26)

----- Final Results -----
               bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase
          [Streptococcus pneumoniae]
 Identities = 133/431 (30%), Positives = 228/431 (52%), Gaps = 20/431 (4%)

Query: 5   IIGIFSLIIIAILAWQGFSFLKHK--EIKLQQAVVEKEIRIAEKTVEVVKRQKTERVLFL 62
           +IGI ++ I  +  +  F   + K  E K++    EK+ +++E   +   RQ    V+
Sbjct: 21  LIGILAISICLLGGFIAFKIYQQKSFEQKIESLKKEKDDQLSEGNQKEHFRQGQAEVIAY 80

Query: 63  EPKGYDKSLSADILKWNQKSFEHKKFYDNQYIILRPQLADSNFANVKKLSIYQILYQKEK 122
            P   +K +S+     NQ   +  +  DN       Q  +S    V   ++ + +Y
Sbjct: 81  YPLQGEKVISSVRELINQDVKDKLESKDNLVFYYTEQ-EESGLKGVVNRNVTKQIYDLVA 139

Query: 123 GSMFQKSSRLLRTYLLDQNKKPFELDELLAHNISGFKAILENIAPGTQLK--EHDSNKEF 180
             + +     L    L ++ +PF LD+L +      + +++ +    + K  E D +++
Sbjct: 140 FKIEETEKTSLGKVHLTEDGQPFTLDQLFSDASKAKEQLIKELTSFIEDKKIEQDQSEQI 199

Query: 181 LKTGRVTD----GLDVKDGKLII---------NDLKLPLDKLYNVIDESYLKSSDLDLVS 227
           +K     D      D KD ++I+         ++  LP+   ++VI  SYL   D  L
Sbjct: 200 VKNFSDQDLSAWNFDYKDSQIILYPSPVVENLEEIALPVSAFFDVIQSSYLLEKDAALYQ 259

Query: 228 NLKAKAPR--VALTFDDGPNEKTTPKALEILKRYNAKATFFVMGQSAVGHTDILQRMHAE 285
           +   K  +  VALTFDDGPN  TTP+ LE L +Y+ KATFFV+G++  G+ D+++R+ +E
Sbjct: 260 SYFDKKHQKVVALTFDDGPNPATTPQVLETLAKYDIKATFFVLGKNVSGNEDLVKRIKSE 319

Query: 286 GHEIGNHTWDHPNLTKLPAEKIKEEIHKTNDLIMKATGQKPVYLRPPYGATNATVKTVTG 345
           GH +GNH+W HP L++L  ++ K++I  T D++ K  G     +RPPYGA    ++
Sbjct: 320 GHVVGNHSWSHPILSQLSLDEAKKQITDTEDVLTKVLGSSSKLMRPPYGAITDDIRNSLD 379

Query: 346 LKEMLWSVDTEDWKNHNTQAMMTNIKKQLRPGGVILMHDIHQTTIDALPTIMDYLTTQGY 405
           L  ++W VD+ DWK+ N  +++T I+ Q+  G ++LMHDIH  T++ALP +++YL  QGY
Sbjct: 380 LSFIMWDVDSLDWKSKNEASILTEIQHQVANGSIVLMHDIHSPTVNALPRVIEYLKNQGY 439

Query: 406 YFVTVGELYST 416
            FVT+ E+ +T
Sbjct: 440 TFVTIPEMLNT 450
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4459> which encodes the amino acid sequence <SEQ ID 4460>. Analysis of this protein sequence reveals the following:

```
     Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.58   Transmembrane    6 -  22 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
!GB:AJ251472 peptidoglycan GlcNAc deacetylase [Strep...     239  4e-62

>GP:CAB96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase
          [Streptococcus pneumoniae]
 Identities = 136/438 (31%), Positives = 230/438 (52%), Gaps = 23/438 (5%)

Query: 3   KLNVILVGLLSILMLSLAI----VFINRWKLNEDSQRIVLAEKKKNTSDLVIKAVKHIKK 58
           K   +L+ L+ IL +S+ +    +    ++     Q+I   +K+K+         +H ++
Sbjct: 13  KTRHVLLALIGILAISICLLGGFIAFKIYQQKSFEQKIESLKKEKDDQLSEGNQKEHFRQ 72

Query: 59  DQKDYYYFSPIK--QADDFFVDNLPVSLYKKKNSDKELILVRPKLQSSHLRSVNTLTISK 116
            Q +   + P++  +      + +   +  K  S   L+    + + S L+ V    ++K
Sbjct: 73  GQAEVIAYYPLQGEKVISSVRELINQDVKDKLESKDNLVFYYTEQEESGLKGVVNRNVTK 132

Query: 117 IVYQKKFFHLAKKSEKVISTYHVTDDLKPFQVKDLVSGHL---ERIQEEVEKKYPDAGFN 173
            +Y    F + +  +  +   H+T+D +PF +  L S      E++ +E+     D
Sbjct: 133 QIYDLVAFKIEETEKTSLGKVHLTEDGQPFTLDQLFSDASKAKEQLIKELTSFIEDKKIE 192

Query: 174 SDKYNGLKESNS---LLSDGFEVKSGNLIFD--------KKLTIPLTTLFDVINPDFLAN 222
            D+   + ++ S   L +  F+ K   +I          +++ +P++  FDVI   +L
Sbjct: 193 QDQSEQIVKNFSDQDLSAWNFDYKDSQIILYPSPVVENLEEIALPVSAFFDVIQSSYLLE 252

Query: 223 SDRAAYDNYRTYKEQHPKKLVALTFDDGPDPTTTPQVLDILAKYQAKGTFFMIGSKVVNN 282
            D A Y +Y   K Q   K+VALTFDDGP+P TTPQVL+ LAKY  K TFF++G  V  N
Sbjct: 253 KDAALYQSYFDKKHQ---KVVALTFDDGPNPATTPQVLETLAKYDIKATFFVLGKNVSGN 309

Query: 283 ENLTKRVSDAGHEIANHTWDHPNLTNLSVSEIQHQVNMTNQAIEKACGKKPRYLRPPYGA 342
           E+L KR+   GH + NH+W HP L+ LS+ E + Q+  T   + K  G   + +RPPYGA
Sbjct: 310 EDLVKRIKSEGHVVGNHSWSHPILSQLSLDEAKKQITDTEDVLTKVLGSSSKLMRPPYGA 369

Query: 343 TNATVQQSSGLTQMLWTVDTRDWENHSTDGIMTNVKNQLQPGGVVLMHDIHQTTINALPT 402
               ++ S  L+ ++W VD+ DW++ +   I+T +++Q+  G +VLMHDIH  T+NALP
Sbjct: 370 ITDDIRNSLDLSFIMWDVDSLDWKSKNEASILTEIQHQVANGSIVLMHDIHSPTVNALPR 429

Query: 403 VMEYLKAEGYECVTVSEL 420
           V+EYLK +GY  VT+ E+
Sbjct: 430 VIEYLKNQGYTFVTIPEM 447
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 169/420 (40%), Positives = 259/420 (61%), Gaps = 12/420 (2%)

Query: 4   LIIGIFSLIIIAILAWQGFSFLKHKEIKLQQAVVEKEIRIAEKTVEVVKRQKTER--VLF 61
           +++G+ S+++++ LA    +  K  E   +  + EK+   ++  ++ VK  K ++    +
Sbjct: 7   ILVGLLSILMLS-LAIVFINRWKLNEDSQRIVLAEKKKNTSDLVIKAVKHIKKDQKDYYY 65

Query: 62  LEPKGYDKSLSADILKWNQKSFEHKKFYDNQYIILRPQLADSNFANVKKLSIYQILYQKE 121
             P         D L     S   KK  D + I++RP+L  S+  +V  L+I +I+YQK+
Sbjct: 66  FSPIKQADDFFVDNLP---VSLYKKKNSDKELILVRPKLQSSHLRSVNTLTISKIVYQKK 122

Query: 122 KGSMFQKSSRLLRTYLLDQNKKPFELDELLAHNISGFKAILENIAPGTQLKEHDSNKEFL 181
              + +KS +++ TY +  + KPF++ +L++ ++   +  +E   P         N
Sbjct: 123 FFHLAKKSEKVISTYHVTDDLKPFQVKDLVSGHLERIQEEVEKKYPDAGFNSDKYNGLKE 182

Query: 182 KTGRVTDGLDVKDGKLIIND-LKLPLDKLYNVIDESYLKSSDLDLVSNL---KAKAPR-- 235
               ++DG +VK G LI +  L +PL  L++VI+  +L +SD     N    K + P+
Sbjct: 183 SNSLLSDGFEVKSGNLIFDKKLTIPLTTLFDVINPDFLANSDRAAYDNYRTYKEQHPKKL 242

Query: 236 VALTFDDGPNEKTTPKALEILKRYNAKATFFVMGQSAVGHTDILQRMHAEGHEIGNHTWD 295
           VALTFDDGP+  TTP+ L+IL +Y AK TFF++G   V + ++ +R+   GHEI NHTWD
Sbjct: 243 VALTFDDGPDPTTTPQVLDILAKYQAKGTFFMIGSKVVNNENLTKRVSDAGHEIANHTWD 302

Query: 296 HPNLTKLPAEKIKEEIHKTNDLIMKATGQKPVYLRPPYGATNATVKTVTGLKEMLWSVDT 355
           HPNLT L   +I+ +++ TN  I KA G+KP YLRPPYGATNATV+  +GL +MLW+VDT
Sbjct: 303 HPNLTNLSVSEIQHQVNMTNQAIEKACGKKPRYLRPPYGATNATVQQSSGLTQMLWTVDT 362
```

```
Query: 356 EDWKNHNTQAMMTNIKKQLRPGGVILMHDIHQTTIDALPTIMDYLTTQGYYFVTVGELYS 415
            DW+NH+T  +MTN+K QL+PGGV+LMHDIHQTTI+ALPT+M+YL  +GY  VTV ELY+
Sbjct: 363 RDWENHSTDGIMTNVKNQLQPGGVVLMHDIHQTTINALPTVMEYLKAEGYECVTVSELYA 422
```

GBS281d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 8-10; MW 71.5kDa) and in Figure 187 (lane 10; MW 71kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 12; MW 46.5kDa) and in Figure 183 (lane 2; MW 46kDa). Purified GBS281d-GST is shown in lane 6 of Figure 237.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1453

A DNA sequence (GBSx1539) was identified in *S.agalactiae* <SEQ ID 4461> which encodes the amino acid sequence <SEQ ID 4462>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4463> which encodes the amino acid sequence <SEQ ID 4464>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2799(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 311/475 (65%), Positives = 389/475 (81%)

Query: 1   MTKEYQNYVNGEWKSSVNQIEILSPIDDSSLGFVPAMTREEVDHAMKAGREALPAWAALT 60
           + K+Y+N VNGEWK S N+I I +P     LG VPAMT+ EVD    + ++AL  W AL+
Sbjct: 1   LAKQYKNLVNGEWKLSENEITIYAPATGEELGSVPAMTQAEVDAVYASAKKALSDWRALS 60

Query: 61  VYERAQYLHKAADIIERDKEEIATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120
             ERA YLHKAADI+ RD E+I  +L+KE++K + A+V+EV+RTA++I YAAEEG+R+
Sbjct: 61  YVERAAYLHKAADILVRDAEKIGAILSKEVAKGHKAAVSEVIRTAEIINYAAEEGLRMEG 120

Query: 121 SADEGGKMDASTGHKLAVIRRQPVGIVLAIAPYNYPVNLSGSKIAPALIGGNVVMFKPPT 180
              EGG  +A++  K+A++RR+PVG+VLAI+P+NYPVNL+GSKIAPALI GNVV  KPPT
Sbjct: 121 EVLEGGSFEAASKKKIAIVRREPVGLVLAISPFNYPVNLAGSKIAPALIAGNVVALKPPT 180

Query: 181 QGSVSGLVLAKAFAEAGLPAGVFNTITGRGSEIGDYIVEHEEVNFINFTGSTPVGKRIGK 240
           QGS+SGL+LA+AFAEAG+PAGVFNTITGRGS IGDYIVEHE V+FINFTGSTP+G+ IGK
Sbjct: 181 QGSISGLLLAEAFAEAGIPAGVFNTITGRGSVIGDYIVEHEAVSFINFTGSTPIGEGIGK 240

Query: 241 LAGMRPIMLELGGKDAGVVLADADLDNAAKQIVAGAYDYSGQRCTAIKRVLVVEEVADEL 300
           LAGMRPIMLELGGKD+ +VL DADL  AAK IVAGA+ YSGQRCTA+KRVLV+++VAD+L
Sbjct: 241 LAGMRPIMLELGGKDSAIVLEDADLALAAKNIVAGAFGYSGQRCTAVKRVLVMDKVADQL 300

Query: 301 AEKISENVAKLSVGDPFDNATVTPVIDDNSADFIESLVVDARQKGAKELNEFKRDGRLLT 360
```

```
           A +I   V KLSVG P D+A +TP+ID ++ADF+E L+ DA  KGA  L  F R+G L++
Sbjct: 301 AAEIKTLVEKLSVGMPEDDADITPLIDTSAADFVEGLIKDATDKGATALTAFNREGNLIS 360

Query: 361 PGLFDHVTLDMKLAWEEPFGPILPIIRVKDAEEAVAIANKSDFGLQSSVFTRDFQKAFDI 420
           P LFDHVT DM+LAWEEPFGP+LPIIRV   EEA+ I+N+S++GLQ+S+FT +F KAF I
Sbjct: 361 PVLFDHVTTDMRLAWEEPFGPVLPIIRVTTVEEAIKISNESEYGLQASIFTTNFPKAFGI 420

Query: 421 ANKLEVGTVHINNKTGRGPDNFPFLGLKGSGAGVQGIRYSIEAMTNVKSIVFDMK 475
           A +LEVGTVH+NNKT RG DNFPFLG K SGAGVQG++YSIEAMT VKS+VFD++
Sbjct: 421 AEQLEVGTVHLNNKTQRGTDNFPFLGAKKSGAGVQGVKYSIEAMTTVKSVVFDIQ 475
```

A related GBS gene <SEQ ID 8815> and protein <SEQ ID 8816> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:    -15.11
GvH: Signal Score (-7.5): 0.17
     Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value:   1.22 threshold:  0.0
   PERIPHERAL  Likelihood =  1.22     187
 modified ALOM score:  -0.74

*** Reasoning Step: 3

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
66.8/82.6% over 474aa
                                                                 Streptococcus mutans
  EGAD|42413| NADP-dependent glyceraldehyde-3-phosphate dehydrogenase Insert characterized
  EGAD|42413|110509   NADP-dependent   glyceraldehyde-3-phosphate   dehydrogenase   Insert
characterized
  SP|Q59931|GAPN_STRMU NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.9)
  (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE) (GLYCERALDEHYDE-3-PHOSPHATE
 DEHYDROGENASE [NADP+]) (TRIOSEPHOSPHATE DEHYDROGENASE). Edit characterized
  GP|642667|gb|AAA91091.1||L38521 NADP-dependent glyceraldehyde-3-phosphate dehydro Insert
characterized

ORF01688(301 - 1725 of 2025)
EGAD|42413|44796(1 - 475 of 475) NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
{Streptococcus    mutans}EGAD|42413|110509    NADP-dependent    glyceraldehyde-3-phosphate
dehydrogenase  {Streptococcus  mutans}SP|Q59931|GAPN_STRMU NADP-DEPENDENT GLYCERALDEHYDE-3-
PHOSPHATE  DEHYDROGENASE  (EC  1.2.1.9)  (NON-PHOSPHORYLATING  GLYCERALDEHYDE  3-PHOSPHATE
DEHYDROGENASE)    (GLYCERALDEHYDE-3-PHOSPHATE    DEHYDROGENASE    [NADP+])    (TRIOSEPHOSPHATE
DEHYDROGENASE).GP|642667|gb|AAA91091.1||L38521   NADP-dependent   glyceraldehyde-3-phosphate
dehydro
%Match = 49.3
%Identity = 66.7  %Similarity = 82.5
Matches = 317  Mismatches = 83  Conservative Sub.s = 75

195       225       255       285       315       345       375       405
*GLKNLYFFIESLDIVKFLRKICQIIEINR*SDRINLLQCKRRFTLTKEYQNYVNGEWKSSVNQIEILSPIDDSSLGFVP
                                             :||:|:|||||||| | |:|:|  |    :|| ||
                                             MTKQYKNYVNGEWKLSENEIKIYEPASGAELGSVP
                                                     10        20        30

435       465       495       525       555       585       615       645
AMTREEVDHAMKAGREALPAWAALTVYERAQYLHKAADIIERDKEEIATVLAKEISKAYNASVTEVVRTADLIRYAAEEG
||: ||||:   : ::| ||| ||:  |||| |||: ||||:|  :|:||::| | ::|:||||||::| ||||||
AMSTEEVDYVYASAKKAQPAWRALSYIERAAYLHKVADILMRDKEKIGAILSKEVAKGYKSAVSEVVRTAEIINYAAEEG
             50        60        70        80        90        100       110
```

```
675       705       735       765       795       825       855       885
IRLSTSADEGGKMDASTGHKLAVIRRQPVGIVLAIAPYNYPVNLSGSKIAPALIGGNVVMFKPPTQGSVSGLVLAKAFAE
:|:     |||  :|::  |:||:||:|||:||||:|:||||||:||||||||| |||: ||||||||:|||:||:||||
LRMEGEVLEGGSFEAASKKKIAVVRREPVGLVLAISPFNYPVNLAGSKIAPALIAGNVIAFKPPTQGSISGLLLAEAFAE
            130       140       150       160       170       180       190

915       945       975       1005      1035      1065      1095      1125
AGLPAGVFNTITGRGSEIGDYIVEHEEVNFINFTGSTPVGKRIGKLAGMRPIMLELGGKDAGVVLADADLDNAAKQIVAG
|||||||||||||||||||||||||: |||||||||| :|:||||:||||||||||||||: :|| ||||:  || |:||
AGLPAGVFNTITGRGSEIGDYIVEHQAVNFINFTGSTGIGERIGKMAGMRPIMLELGGKDSAIVLEDADLELTAKNIIAG
            210       220       230       240       250       260       270

1155      1185      1215      1245      1275      1305      1335      1365
AYDYSGQRCTAIKRVLVVEEVADELAEKISENVAKLSVGXPFDNATVTPVIDDNSADFIESLVVDARQKGAKELNEFKRD
|: ||||||||:|||||:| ||||| ||| | |  |::| | |:| :||:||  |||::| |: ||  |||  | | ||:
AFGYSGQRCTAVKRVLVMESVADELVEKIREKVLALTIGNPEDDADITPLIDTKSADYVEGLINDANDKGATALTEIKRE
            290       300       310       320       330       340       350

1395      1425      1455      1485      1515      1545      1575      1605
GRLLTPGLFDHVTLDMKLAWEEPFGPILPIIRVKDAEEAVAIANKXDFGLQSSVFTRDFQKAFDIANKLEVGTVHINNKT
| |: | ||| || ||:|||||||||:||||||   |||: |:|| ::|||:|:|| || :|| || :||||||||||||
GNLICPILFDKVTTDMRLAWEEPFGPVLPIIRVTSVEEAIEISNKSEYGLQASIFTNDFPRAFGIAEQLEVGTVHINNKT
            370       380       390       400       410       420       430

1635      1665      1695      1725      1755      1785      1815      1845
GRGPDNFPFLGLKGSGAGVQGIRYSIEAMTNVKSIVFDMK*T*NDSTIVS*VVL*TSFTLKIKNYIIF*SGFIFVI*LS*
 || ||||||| | ||||:||::||||||| |||:|||:|
QRGTDNFPFLGAKKSGAGIQGVKYSIEAMTTVKSVVFDIK
            450       460       470
```

SEQ ID 8816 (GBS127) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 10; MW 55.9kDa).

GBS127-His was purified as shown in Figure 200, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1454

A DNA sequence (GBSx1540) was identified in *S.agalactiae* <SEQ ID 4465> which encodes the amino acid sequence <SEQ ID 4466>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  427 - 443 ( 427 - 443)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA78049 GB:AB027569 phosphoenolpyruvate-protein
           phosphotransferase [Streptococcus bovis]
 Identities = 534/577 (92%), Positives = 559/577 (96%)

Query: 1    MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTNAEEARLDVALQASQDELSVIRE 60
            MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDT+AEEARLD AL+ASQDELS+IRE
Sbjct: 1    MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTSAEEARLDAALKASQDELSIIRE 60

Query: 61   KAVESLGEEAAAVFDAHLMVLSDPEMINQIKETIRAKQVNAETGLKEVTDMFITIFEGME 120
            KAVE+LGEEAAAVFDAHLMVL+DPEMI+QIKETIRAKQ NAE GLKEVTDMFITIFEGME
Sbjct: 61   KAVETLGEEAAAVFDAHLMVLADPEMISQIKETIRAKQTNAEAGLKEVTDMFITIFEGME 120
```

```
Query: 121 DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180
           DNPYMQERAADIRDVAKRVLAHLLG KLPNPATI+EESIVIAHDLTPSDTAQLNKQFVKA
Sbjct: 121 DNPYMQERAADIRDVAKRVLAHLLGAKLPNPATIDEESIVIAHDLTPSDTAQLNKQFVKA 180

Query: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITERVQDGQLIAVNGITGEVIIEPTEAQISA 240
           FVTNIGGRTSHSAIMARTLEIAAVLGTNDIT RV+DG ++AVNGITGEVII PT+ Q++
Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITSRVKDGDIVAVNGITGEVIINPTDEQVAE 240

Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAANIGTPKDVEGVNENGAEAVGLYRTE 300
           FKAAGEAYAKQKAEWALLKDA+TVTADGKHFELAANIGTPKDVEGVN NGAEAVGLYRTE
Sbjct: 241 FKAAGEAYAKQKAEWALLKDAKTVTADGKHFELAANIGTPKDVEGVNANGAEAVGLYRTE 300

Query: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360
           FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPY DLPKEMNPFLGFR
Sbjct: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYLDLPKEMNPFLGFR 360

Query: 361 ALRISISETGDAMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEEKANLLAD 420
           ALRISISETG+AMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIF+EEKANL A+
Sbjct: 361 ALRISISETGNAMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFDEEKANLKAE 420

Query: 421 GVAVAEGIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQVSYLYQP 480
           GVAV++ I+VGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQVSYLYQP
Sbjct: 421 GVAVSDDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQVSYLYQP 480

Query: 481 YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQTAVPLLVGMGLDEFSMSATSVLRTRSL 540
           YNPSILRLINNVIKAAHAEGKW GMCGEMAGDQ AVPLLV MGLDEFSMSATS+LRTRSL
Sbjct: 481 YNPSILRLINNVIKAAHAEGKWVGMCGEMAGDQKAVPLLVEMGLDEFSMSATSILRTRSL 540

Query: 541 MKKLDTAKMEEYANRALSECSTMEEVIELQKEYVDFD 577
           MKKLDTAKM+EYANRAL+ECSTMEEV+EL KEYV+ D
Sbjct: 541 MKKLDTAKMQEYANRALTECSTMEEVLELSKEYVNVD 577
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4467> which encodes the amino acid sequence <SEQ ID 4468>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0875(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 540/577 (93%), Positives = 561/577 (96%)

Query: 1   MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTNAEEARLDVALQASQDELSVIRE 60
            MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTV DTNAEEARLDVALQA+QDELSVIRE
Sbjct: 1   MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVADTNAEEARLDVALQAAQDELSVIRE 60

Query: 61  KAVESLGEEAAAVFDAHLMVLSDPEMINQIKETIRAKQVNAETGLKEVTDMFITIFEGME 120
             AVESLGEEAAAVFDAHLMVL+DPEMI+Q+KETIRAKQ NAETGLKEVTDMFITIFEGME
Sbjct: 61  NAVESLGEEAAAVFDAHLMVLADPEMISQVKETIRAKQTNAETGLKEVTDMFITIFEGME 120

Query: 121 DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180
            DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA
Sbjct: 121 DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180

Query: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITERVQDGQLIAVNGITGEVIIEPTEAQISA 240
            FVTNIGGRTSHSAIMARTLEIAAVLGTNDIT+RV+DG +IAVNGITGEVII+P+E Q+ A
Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITKRVKDGDVIAVNGITGEVIIDPSEDQVLA 240

Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAANIGTPKDVEGVNENGAEAVGLYRTE 300
            FK AG AYAKQKAEW+LLKDA T TADGKHFELAANIGTPKDVEGVN+NGAEAVGLYRTE
Sbjct: 241 FKEAGAAYAKQKAEWSLLKDAHTETADGKHFELAANIGTPKDVEGVNDNGAEAVGLYRTE 300
```

```
Query: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360
           FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR
Sbjct: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360

Query: 361 ALRISISETGDAMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEEKANLLAD 420
           ALRISISETGDAMFRTQ+RALLRASVHGQLRIMFPMVALLKEFRAAKA+F+EEKANLLA+
Sbjct: 361 ALRISISETGDAMFRTQMRALLRASVHGQLRIMFPMVALLKEFRAAKAVFDEEKANLLAE 420

Query: 421 GVAVAEGIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQVSYLYQP 480
           GVAVA+ I+VGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQVSYLYQP
Sbjct: 421 GVAVADDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMAADRMNEQVSYLYQP 480

Query: 481 YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQTAVPLLVGMGLDEFSMSATSVLRTRSL 540
           YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQ AVPLLVGMGLDEFSMSATSVLRTRSL
Sbjct: 481 YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQQAVPLLVGMGLDEFSMSATSVLRTRSL 540

Query: 541 MKKLDTAKMEEYANRALSECSTMEEVIELQKEYVDFD 577
           MKKLD+AKMEEYANRAL+ECST EEV+EL KEYV  D
Sbjct: 541 MKKLDSAKMEEYANRALTECSTAEEVLELSKEYVSED 577
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1455

A DNA sequence (GBSx1541) was identified in *S.agalactiae* <SEQ ID 4469> which encodes the amino acid sequence <SEQ ID 4470>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein is similar to a protein from *S.bovis*:

```
>GP:BAA78048 GB:AB027569 histidine containing protein [Streptococcus bovis]
 Identities = 86/87 (98%), Positives = 87/87 (99%)

Query: 1  MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD 60
          MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD
Sbjct: 1  MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD 60

Query: 61 VTISAEGADADDAIAAIEETMTKEGLA 87
          VTISAEGADADDA+AAIEETMTKEGLA
Sbjct: 61 VTISAEGADADDALAAIEETMTKEGLA 87
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4471> which encodes the amino acid sequence <SEQ ID 4472>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 86/87 (98%), Positives = 87/87 (99%)
```

```
Query: 1  MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD 60
          MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD
Sbjct: 1  MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD 60

Query: 61 VTISAEGADADDAIAAIEETMTKEGLA 87
          VTISAEGADA+DAIAAIEETMTKEGLA
Sbjct: 61 VTISAEGADAEDAIAAIEETMTKEGLA 87
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1456

A DNA sequence (GBSx1542) was identified in *S.agalactiae* <SEQ ID 4473> which encodes the amino acid sequence <SEQ ID 4474>. This protein is predicted to be glutaredoxin-like protein nrdh (b2673). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4532(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA63372 GB:X92690 glutaredoxin-like protein [Lactococcus
          lactis]
 Identities = 42/70 (60%), Positives = 53/70 (75%)

Query: 4  ITVFSKNNCMQCKMTKKFLDQHGADFEEINIDEKPEKIEYVKNLGFSAAPVIEAGNVVFS 63
          +TV+SKNNCMQCKM KK+L +H   F EINIDE+PE +E V  +GF AAPVI   +  FS
Sbjct: 2  VTVYSKNNCMQCKMVKKWLSEHEIAFNEINIDEQPEFVEKVIEMGFRAAPVITKDDFAFS 61

Query: 64 GFQPSKLKEL 73
          GF+PS+L +L
Sbjct: 62 GFRPSELAKL 71
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4475> which encodes the amino acid sequence <SEQ ID 4476>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4606(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 56/71 (78%), Positives = 68/71 (94%)

Query: 4  ITVFSKNNCMQCKMTKKFLDQHGADFEEINIDEKPEKIEYVKNLGFSAAPVIEAGNVVFS 63
          ITV+SKNNCMQCKMTKKFL+QHG +F+EINIDE PEK++YVK+LGF++APVIEA N+VFS
Sbjct: 13 ITVYSKNNCMQCKMTKKFLEQHGVNFQEINIDEHPEKVDYVKSLGFTSAPVIEADNLVFS 72

Query: 64 GFQPSKLKELV 74
          GFQP+KLKEL+
Sbjct: 73 GFQPAKLKELI 83
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1457

A DNA sequence (GBSx1543) was identified in *S.agalactiae* <SEQ ID 4477> which encodes the amino acid sequence <SEQ ID 4478>. This protein is predicted to be ribonucleotide reductase subunit R1E (nrdE). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3676(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD41036 GB:AF112535 ribonucleotide reductase alpha-chain
           [Corynebacterium glutamicum]
 Identities = 366/701 (52%), Positives = 488/701 (69%), Gaps = 19/701 (2%)

Query: 23  NGQIPLHKDKEALTAFFKENVQPNSKAFDSITDKIAYLLKYDYLEEAFLNKYRPEFIEEL 82
            NG+I   KD+EA   +F ++V  N+  F ++ +KI YL++  Y +   L+KY  +FI++L
Sbjct: 22  NGKIQFEKDREAANQYFLQHVNQNTVFFHNLQEKIDYLVENKYYDPIVLDKYDFQFIKDL 81

Query: 83  STKLFDKKFRFKSFMAAYKFYQQYALKTNDGEYYLESIEDRVLFNALYFADGDEELATDL 142
              + +   KFRF+SF+ AYK+Y  Y LKT DG  YLE  EDRV   AL  ADGD  LA +L
Sbjct: 82  FKRAYGFKFRFQSFLGAYKYYTSYTLKTFDGRRYLERFEDRVCMVALTLADGDRALAENL 141

Query: 143 ALEMISQRYQPATPSFLNAGRSRRGELVSCFLIQVTDDMNAIGRSINSALQLSRIGGGVG 202
              E++S R+QPATP+FLN+G+++RGE VSCFL+++ D+M +IGRSINSALQLS+ GGGV
Sbjct: 142 VDEIMSGRFQPATPTFLNSGKAQRGEPVSCFLLRIEDNMESIGRSINSALQLSKRGGGVA 201

Query: 203 ISLSNLREAGAPIKGFAGAASGVVPVMKLFEDSFSYSNQLGQRQGAGVVYLDVFHPDIIS 262
            + LSNLREAGAPIK     +SGV+PVMKL ED+FSY+NQLG RQGAG VYL+  HPDI+S
Sbjct: 202 LLLSNLREAGAPIKKIENQSSGVIPVMKLLEDAFSYANQLGARQGAGAVYLNAHHPDILS 261

Query: 263 FLSTKKENADEKVRVKTLSLGITVPDKFYELARNNQEMYLFSPYSIEREYGVPFSYIDIT 322
            FL TK+ENADEK+R+KTLSLG+ +PD  +ELA+ N +MYLFSPY +ER YG PF+ + IT
Sbjct: 262 FLDTKRENADEKIRIKTLSLGVVIPDITFELAKRNDDMYLFSPYDVERIYGKPFADVSIT 321

Query: 323 EKYDELVANPNITKTKINARDLETEISKLQQESGYPYIINIDTANRTNPVDGKIIMSNLC 382
            E YDE+V +  I KTKINAR     ++++Q ESGYPYI+  DT N +NP++G+I  SNLC
Sbjct: 322 EHYDEMVDDDRIRKTKINARQFFQTLAEIQFESGYPYIMYEDTVNASNPIEGRITHSNLC 381

Query: 383 SEILQVQKPSLINDAQEYLEMGTDISCNLGSTNVLNMMTSPDFGKSIKTMTRALTFVTDS 442
            SEILQV  PS  ND    Y E+G DISCNLGS NV   M SP+F K+I+T  R LT V++
Sbjct: 382 SEILQVSTPSEFNDDLTYAEVGEDISCNLGSLNVAMAMDSPNFEKTIETAIRGLTAVSEQ 441

Query: 443 SNIEAVPTIKNGNAQAHTFGLGAMGLHSYLAKNHIEYGSPESIEFTDIYFMLMNYWTLVE 502
            ++I++VP+I+ GN  AH  GLG M LH Y  + H+ YGS E+++FT+ YF  + Y  L
Sbjct: 442 TSIDSVPSIRKGNEAAHAIGLGQMNLHGYFGREHMHYGSEEALDFTNAYFAAVLYQCLRA 501

Query: 503 SNNIARERQTTFVGFEKSKYADGTYFDKYVSGKFVPQSDKVKSLFA--NHFIPEAKDWEN 560
            SN IA ER   F  FE SKYA G YFD + +  F P+SDKVK LFA  N   P  +DW
Sbjct: 502 SNKIATERGERFKNFENSKYATGEYFDDFDANDFAPKSDKVKELFAKSNIHTPTVEDWAA 561

Query: 561 LRYAVMKDGLYHQNRLAVAPNGSISYINDCSASIHPITQRIEERQEKKIGKIYYPANGLA 620
            L+  VM+ GL+++N  AV P GSISYIN+ ++SIHPI  +IE R+E KIG++YYPA  +
Sbjct: 562 LKADVMEHGLFNRNLQAVPPTGSISYINNSTSSIHPIASKIEIRKEGKIGRVYYPAPHMD 621

Query: 621 TDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSMTLFLRSELPKELYEWKTESKQTTRD 680
             D + Y+  AY++   K+ID YA AT++VDQGLS+TLF +                TTRD
Sbjct: 622 NDNLEYFEDAYEIGYEKIIDTYAVATKYVDQGLSLTLFFK-------------DTATTRD 668
```

```
Query: 681 LSILRNYAFNKGVKSIYYI--RTFTDDGSEVGANQCESCVI 719
           ++  + YA+ KG+K++YYI  R    +G+EV  + C SC++
Sbjct: 669 INRAQIYAWRKGIKTLYYIRLRQVALEGTEV--DGCVSCML 707
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4479> which encodes the amino acid sequence <SEQ ID 4480>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 628/719 (87%), Positives = 682/719 (94%)

Query: 1   MSLKNIGDVSYFRLNNEINRPVNGQIPLHKDKEALTAFFKENVQPNSKAFDSITDKIAYL 60
           MSLK++GD+SYFRLNNEINRPVNG+IPLHKDKEAL AF  ENV PN+ +F SIT+KI YL
Sbjct: 1   MSLKDLGDISYFRLNNEINRPVNGKIPLHKDKEALKAFSAENVLPNTMSFTSITEKIEYL 60

Query: 61  LKYDYLEEAFLNKYRPEFIEELSTKLFDKKFRFKSFMAAYKFYQQYALKTNDGEYYLESI 120
           +  DY+E AF+ KYRPEFI EL + +  + FRFKSFMAAYKFYQQYALKTNDGE+YLE++
Sbjct: 61  ISNDYIESAFIQKYRPEFITELDSIIKSENFRFKSFMAAYKFYQQYALKTNDGEHYLENL 120

Query: 121 EDRVLFNALYFADGDEELATDLALEMISQRYQPATPSFLNAGRSRRGELVSCFLIQVTDD 180
           EDRVLFNALYFADG E+LA DLA+EMI+QRYQPATPSFLNAGRSRRGELVSCFLIQVTDD
Sbjct: 121 EDRVLFNALYFADGQEDLAKDLAVEMINQRYQPATPSFLNAGRSRRGELVSCFLIQVTDD 180

Query: 181 MNAIGRSINSALQLSRIGGGVGISLSNLREAGAPIKGFAGAASGVVPVMKLFEDSFSYSN 240
           MN+IGRSINSALQLSRIGGGVGI+LSNLREAGAPIKG+AGAASGVVPVMKLFEDSFSYSN
Sbjct: 181 MNSIGRSINSALQLSRIGGGVGITLSNLREAGAPIKGYAGAASGVVPVMKLFEDSFSYSN 240

Query: 241 QLGQRQGAGVVYLDVFHPDIISFLSTKKENADEKVRVKTLSLGITVPDKFYELARNNQEM 300
           QLGQRQGAGVVYL+VFHPDII+FLSTKKENADEKVRVKTLSLGITVPDKFYELAR N++M
Sbjct: 241 QLGQRQGAGVVYLNVFHPDIIAFLSTKKENADEKVRVKTLSLGITVPDKFYELARKNEDM 300

Query: 301 YLFSPYSIEREYGVPFSYIDITEKYDELVANPNITKTKINARDLETEISKLQQESGYPYI 360
           YLFSPY++E+EYG+PF+Y+DIT  YDELVANP ITKTKI ARDLETEISKLQQESGYPYI
Sbjct: 301 YLFSPYNVEKEYGIPFNYLDITNMYDELVANPKITKTKIKARDLETEISKLQQESGYPYI 360

Query: 361 INIDTANRTNPVDGKIIMSNLCSEILQVQKPSLINDAQEYLEMGTDISCNLGSTNVLNMM 420
           INIDTAN+ NP+DGKIIMSNLCSEILQVQ PSLINDAQE++EMGTDISCNLGSTN+LNMM
Sbjct: 361 INIDTANKANPIDGKIIMSNLCSEILQVQTPSLINDAQEFVEMGTDISCNLGSTNILNMM 420

Query: 421 TSPDFGKSIKTMTRALTFVTDSSNIEAVPTIKNGNAQAHTFGLGAMGLHSYLAKNHIEYG 480
           TSPDFG+SIKTMTRALTFVTDSS+IEAVPTIK+GN+QAHTFGLGAMGLHSYLA++HIEYG
Sbjct: 421 TSPDFGRSIKTMTRALTFVTDSSSIEAVPTIKHGNSQAHTFGLGAMGLHSYLAQHHIEYG 480

Query: 481 SPESIEFTDIYFMLMNYWTLVESNNIARERQTTFVGFEKSKYADGTYFDKYVSGKFVPQS 540
           SPESIEFTDIYFML+NYWTLVESNNIARERQTTFVGFE SKYA+G+YFDKYV+G FVP+S
Sbjct: 481 SPESIEFTDIYFMLLNYWTLVESNNIARERQTTFVGFENSKYANGSYFDKYVTGHFVPKS 540

Query: 541 DKVKSLFANHFIPEAKDWENLRYAVMKDGLYHQNRLAVAPNGSISYINDCSASIHPITQR 600
           D VK LF +HFIP+A DWE LR AV KDGLYHQNRLAVAPNGSISYINDCSASIHPITQR
Sbjct: 541 DLVKDLFKDHFIPQASDWEALRDAVQKDGLYHQNRLAVAPNGSISYINDCSASIHPITQR 600

Query: 601 IEERQEKKIGKIYYPANGLATDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSMTLFLR 660
           IEERQEKKIGKIYYPANGL+TDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLS+TLFLR
Sbjct: 601 IEERQEKKIGKIYYPANGLSTDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSLTLFLR 660

Query: 661 SELPKELYEWKTESKQTTRDLSILRNYAFNKGVKSIYYIRTFTDDGSEVGANQCESCVI 719
           SELP ELYEWKT+SKQTTRDLSILRNYAFNKG+KSIYYIRTFTDDG EVGANQCESCVI
Sbjct: 661 SELPMELYEWKTQSKQTTRDLSILRNYAFNKGIKSIYYIRTFTDDGEEVGANQCESCVI 719
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1458

A DNA sequence (GBSx1544) was identified in *S.agalactiae* <SEQ ID 4481> which encodes the amino acid sequence <SEQ ID 4482>. This protein is predicted to be ribonucleotide reductase subunit R2F (nrdB). Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4583(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9753> which encodes amino acid sequence <SEQ ID 9754> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC14561 GB:AF050168 ribonucleoside diphosphate reductase small
            subunit [Corynebacterium ammoniagenes]
 Identities = 166/313 (53%), Positives = 215/313 (68%), Gaps = 1/313 (0%)

Query: 10  EAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFGGLTLL 69
           +AINWN I D  D   W++LT  FWL  +IP+SND+  W K++ QE+     +VF GLTLL
Sbjct: 17  KAINWNVIPDEKDLEVWDRLTGNFWLPEKIPVSNDIQSWNKMTPQEQLATMRVFTGLTLL 76

Query: 70  DTMQSETGVEAIRADVRTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEIFEWTN 129
           DT+Q   G  ++  DV T HEE V  NI FMESVHAKSYS+IF TL +  +I E F W+
Sbjct: 77  DTIQGTVGAISLLPDVETMHEEGVYTNIAFMESVHAKSYSNIFMTLASTPQINEAFRWSE 136

Query: 130 NNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFFTPLYYLGNNKLANVAEIIK 189
            NE LQ KA+II   Y   + L+KKVAST LE+FLFYSGF+ P+Y     KL N A+II+
Sbjct: 137 ENENLQRKAKIIMSYYNGDDPLKKKVASTLLESFLFYSGFYLPMYLSSRAKLTNTADIIR 196

Query: 190 LIIRDESVHGTYIGYKFQLGFNELPEDEQENFRDWMYDLLYQLYENEEKYTKTLYDGVGW 249
           LIIRDESVHG YIGYK+Q G  +L E EQE ++ + +DL+Y LYENE +YT+ +YD +GW
Sbjct: 197 LIIRDESVHGYYIGYKYQQGVKKLSEAEQEEYKAYTFDLMYDLYENEIEYTEDIYDDLGW 256

Query: 250 TEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIVMNGIS-TGTSNHDFFSQVGNGYLL 308
           TE+V  FLRYNANKAL NLG + LFP     V+P +++ +S     NHDFFS  G+ Y++
Sbjct: 257 TEDVKRFLRYNANKALNNLGYEGLFPTDETKVSPAILSSLSPNADENHDFFSGSGSSYVI 316

Query: 309 GSVEAMHDDDYNY 321
           G  E   DDD+++
Sbjct: 317 GKAEDTTDDDWDF 329
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4483> which encodes the amino acid sequence <SEQ ID 4484>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4583(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 315/319 (98%), Positives = 316/319 (98%)

Query: 5   MTTYYEAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFG 64
            MTTYYEAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLS QEKDLVGKVFG
Sbjct: 1   MTTYYEAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLSLQEKDLVGKVFG 60

Query: 65  GLTLLDTMQSETGVEAIRADVRTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEI 124
            GLTLLDTMQSETGVEAIRADVRTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTK EIEEI
Sbjct: 61  GLTLLDTMQSETGVEAIRADVRTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKKEIEEI 120

Query: 125 FEWTNNNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFFTPLYYLGNNKLANV 184
            FEWTNNNEFLQEKARIINDIYANG+ALQKKVASTYLETFLFYSGFFTPLYYLGNNKLANV
Sbjct: 121 FEWTNNNEFLQEKARIINDIYANGDALQKKVASTYLETFLFYSGFFTPLYYLGNNKLANV 180

Query: 185 AEIIKLIIRDESVHGTYIGYKFQLGFNELPEDEQENFRDWMYDLLYQLYENEEKYTKTLY 244
            AEIIKLIIRDESVHGTYIGYKFQLGFNELPEDEQENFRDWMYDLLYQLYENEEKYTKTLY
Sbjct: 181 AEIIKLIIRDESVHGTYIGYKFQLGFNELPEDEQENFRDWMYDLLYQLYENEEKYTKTLY 240

Query: 245 DGVGWTEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIVMNGISTGTSNHDFFSQVGN 304
            DGVGWTEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIVMNGISTGTSNHDFFSQVGN
Sbjct: 241 DGVGWTEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIVMNGISTGTSNHDFFSQVGN 300

Query: 305 GYLLGSVEAMHDDDYNYGL 323
            GYLLGSVEAM DDDYNYGL
Sbjct: 301 GYLLGSVEAMSDDDYNYGL 319
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1459

A DNA sequence (GBSx1545) was identified in *S.agalactiae* <SEQ ID 4485> which encodes the amino acid sequence <SEQ ID 4486>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane   50 -  66 (  50 -  66)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1460

A DNA sequence (GBSx1546) was identified in *S.agalactiae* <SEQ ID 4487> which encodes the amino acid sequence <SEQ ID 4488>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.38   Transmembrane  176 - 192 ( 168 - 201)
   INTEGRAL    Likelihood = -4.57   Transmembrane   25 -  41 (  22 -  42)
   INTEGRAL    Likelihood = -3.88   Transmembrane   94 - 110 (  94 - 112)
   INTEGRAL    Likelihood = -1.49   Transmembrane   70 -  86 (  70 -  86)
   INTEGRAL    Likelihood = -1.01   Transmembrane  128 - 144 ( 128 - 144)

----- Final Results -----
```

```
          bacterial membrane --- Certainty=0.6753(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9751> which encodes amino acid sequence <SEQ ID 9752> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/184 (29%), Positives = 98/184 (52%), Gaps = 4/184 (2%)

Query: 16  MSKNNNTTCLIETAIFAALAMALSMIP----DFASWFTPSFGAIPLILFALRRGTKYGLF 71
           M+++     LIE AI  A A+ L ++           + S   IP+ L + R G K GL
Sbjct: 1   MNQSKQLVRLIEIAIMTAAAVILDIVSGMFLSMPQGGSVSIMMIPIFLISFRWGVKAGLT 60

Query: 72  AGLIWGLLHFVLSKVYYLSLSQVFIEYILAFISMGLAGVFSAKFKDALSSSSKTKALSLA 131
            GL+ GL+   +  ++     Q+ ++YI+AF ++G++G F++  + A  S +K K +
Sbjct: 61  TGLLTGLVQIAIGNLFAQHPVQLLLDYIVAFAAIGISGCFASSVRKAAVSKTKGKLIVSV 120

Query: 132 LSGAILATLVRYVWHYIAGVIFWASYAPKGMSATLYSLSVNGTAGLLTLFFVVISIIILV 191
           +S   + +L+RY  H I+G +F+ S+APKG    +YSL+ N T  + +     I + +L
Sbjct: 121 VSAVFIGSLLRYAAHVISGAVFFGSFAPKGTPVWIYSLTYNATYMVPSFIICAIVLCLLF 180

Query: 192 ISYP 195
           ++ P
Sbjct: 181 MTAP 184
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4489> which encodes the amino acid sequence <SEQ ID 4490>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.34    Transmembrane  162 - 178 ( 156 - 183)
   INTEGRAL    Likelihood = -9.34    Transmembrane  110 - 126 ( 107 - 130)
   INTEGRAL    Likelihood = -1.22    Transmembrane   55 -  71 (  55 -  71)

----- Final Results -----
               bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/189 (29%), Positives = 100/189 (52%), Gaps = 10/189 (5%)

Query: 1   MSPNTNVKYLIEAAIFAALAMTLSFIPDFAGWF--SPSYGAIALV-----IFSLRRGLKY 53
           M+ +  +  LIE AI  A A+ L  +   +G F   P  G+++++     + S R G+K
Sbjct: 1   MNQSKQLVRLIEIAIMTAAAVILDIV---SGMFLSMPQGGSVSIMMIPIFLISFRWGVKA 57

Query: 54  GMLAGLIWGLLHFVLGKVYYLSMSQVFIEYILAFTSMGLAGSFSDSLIKTLRRQQTFFAV 113
           G+  GL+ GL+   +G ++     Q+ ++YI+AF ++G++G F+ S+ K    +     +
Sbjct: 58  GLTTGLLTGLVQIAIGNLFAQHPVQLLLDYIVAFAAIGISGCFASSVRKAAVSKTKGKLI 117

Query: 114 FLAIMASLLAVTVRYLWHFLAGIIFWGSYAPKGMSAVWYSFSVNGTAGVLTFLITCLALM 173
              + A  +   +RY  H ++G +F+GS+APKG     YS + N T  V +F+I  + L
Sbjct: 118 VSVVSAVFIGSLLRYAAHVISGAVFFGSFAPKGTPVWIYSLTYNATYMVPSFIICAIVLC 177

Query: 174 IALPIHPQL 182
           +     P+L
Sbjct: 178 LLFMTAPRL 186
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 116/186 (62%), Positives = 138/186 (73%)
```

```
Query: 16   MSKNNNTTCLIETAIFAALAMALSMIPDFASWFTPSFGAIPLILFALRRGTKYGLFAGLI 75
            MS N N   LIE AIFAALAM LS IPDFA WF+PS+GAI L++F+LRRG KYG+ AGLI
Sbjct: 1    MSPNTNVKYLIEAAIFAALAMTLSFIPDFAGWFSPSYGAIALVIFSLRRGLKYGMLAGLI 60

Query: 76  WGLLHFVLSKVYYLSLSQVFIEYILAFISMGLAGVFSAKFKDALSSSSKTKALSLALSGA 135
            WGLLHFVL KVYYLS+SQVFIEYILAF SMGLAG FS      L       A+ LA+  +
Sbjct: 61  WGLLHFVLGKVYYLSMSQVFIEYILAFTSMGLAGSFSDSLIKTLRRQQTFFAVFLAIMAS 120

Query: 136 ILATLVRYVWHYIAGVIFWASYAPKGMSATLYSLSVNGTAGLLTLFFVVISIIILVISYP 195
            +LA  VRY+WH++AG+IFW SYAPKGMSA  YS SVNGTAG+LT     ++++I +  +P
Sbjct: 121 LLAVTVRYLWHFLAGIIFWGSYAPKGMSAVWYSFSVNGTAGVLTFLITCLALMIALPIHP 180

Query: 196 SFFLPK 201
              F PK
Sbjct: 181 QLFDPK 186
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1461

A DNA sequence (GBSx1547) was identified in *S.agalactiae* <SEQ ID 4491> which encodes the amino acid sequence <SEQ ID 4492>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.43   Transmembrane  206 - 222 ( 199 - 223)
   INTEGRAL    Likelihood = -6.64   Transmembrane   24 -  40 (  19 -  42)
   INTEGRAL    Likelihood = -6.58   Transmembrane   61 -  77 (  51 -  78)
   INTEGRAL    Likelihood = -6.58   Transmembrane  134 - 150 ( 132 - 154)
   INTEGRAL    Likelihood = -4.62   Transmembrane  226 - 242 ( 224 - 245)
   INTEGRAL    Likelihood = -3.72   Transmembrane  107 - 123 ( 106 - 125)

----- Final Results -----
               bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9749> which encodes amino acid sequence <SEQ ID 9750> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4493> which encodes the amino acid sequence <SEQ ID 4494>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.46   Transmembrane  134 - 150 ( 131 - 159)
   INTEGRAL    Likelihood = -7.59   Transmembrane  107 - 123 ( 103 - 128)
   INTEGRAL    Likelihood = -7.48   Transmembrane  225 - 241 ( 213 - 248)
   INTEGRAL    Likelihood = -7.22   Transmembrane  205 - 221 ( 199 - 224)
   INTEGRAL    Likelihood = -3.56   Transmembrane   50 -  66 (  50 -  73)
   INTEGRAL    Likelihood = -1.28   Transmembrane   16 -  32 (  16 -  33)

----- Final Results -----
               bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 82/253 (32%), Positives = 149/253 (58%), Gaps = 5/253 (1%)

Query: 6   IKQSDTTFVRIIKSLLIGGFIGAILGSVGALFIIF--GQDKYLSEI--NIVQYFLWVSRI 61
           +K+   +F+R++K  L+    G I+G +   F+ +  G+  +L+ +  +++   + ++R+
Sbjct: 1   MKKKKNSFLRLLKMSLLSSLAGGIIGGMVGAFLGYHGGRLDHLTFLKDDVINLIILLNRL 60

Query: 62  VVIITALFSLIYLYQIQKYQKVFFNVDESQ-SEEIYRQINLRHSYGMTFVSISIVLSIVN 120
           VV+     S ++L Q++K   V+  ++E   SE  YRQ+N +H+Y M  ++++ +LS+ N
Sbjct: 61  VVVTDLTLSFVFLTQLKKETAVYNTIEEDDISENGYRQLNKKHAYTMLLIAVASILSMCN 120

Query: 121 TLFNYKLNIFDDSVTLVIPIYDLSLLFVLLGLHIYFLKVYRNIRGIKMTVAPTLKELKNN 180
            L    L        L IP+ D+ LL +++      +K Y  IRG  +   P LKELK+N
Sbjct: 121 VLLGLTLTNDSQHAMLAIPLLDILLLLMVIPFQALAMKRYNAIRGTDVPYFPNLKELKHN 180

Query: 181 VLQLDEAELESNYKMCFDIVMNLSGFIFPTIYFVLFFISFVFQKVEIVAIIITTSIHIYI 240
           ++ LDEAEL++ +K  F+ V++L+G I P++Y +LFF+     +VE+ AI++   I +Y+
Sbjct: 181 IMALDEAELQAYHKTSFESVLSLNGVIIPSLYVILFFVYLFTGQVELTAILVLVLIQLYL 240

Query: 241 LIKSLKAARHFYR 253
           L+KS    R FYR
Sbjct: 241 LVKSATMTRQFYR 253
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1462

A DNA sequence (GBSx1548) was identified in *S.agalactiae* <SEQ ID 4495> which encodes the amino acid sequence <SEQ ID 4496>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5172(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1463

A DNA sequence (GBSx1549) was identified in *S.agalactiae* <SEQ ID 4497> which encodes the amino acid sequence <SEQ ID 4498>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2059(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC76650 GB:AE000440
           UDP-D-glucose:(galactosyl)lipopolysaccharide
```

```
             glucosyltransferase [Escherichia coli K12]
  Identities = 70/256 (27%), Positives = 121/256 (46%), Gaps = 14/256 (5%)

Query: 1   MNLLFSIDDMYVDHFKVMLYSLVRQTKNRKLEIYVLQKT----LLKRHTELIQYTQNLEV 56
           +N+ + +D  Y+D   V + S+V   ++  L+ Y++         ++  +L +  Q
Sbjct: 28  LNVAYGVDANYLDGVGVSITSIVLNNRHINLDFYIIADVYNDGFFQKIAKLAEQNQLRIT 87

Query: 57  GYHPIIVGTEVFAQAPTTDRYPDTIYYRLLAHKFLPETLDRILYLDADMLCLNDFSSLYD 116
            Y    + T+     P T  +   +Y+RL A + L  TLDR+LYLDAD++C  D S L
Sbjct: 88  LYR---INTDKLQCLPCTQVWSRAMYFRLFAFQLLGLTLDRLLYLDADVVCKGDISQLLH 144

Query: 117 MELGDQLYAAASHNTDGKFLDYVNKLRLKNVELESSYFNTGVLLMNLPAIRKVVHQQTIL 176
           + L     A A+   D + +      RL + EL   YFN+GV+ ++L         +  L
Sbjct: 145 LGLNG---AVAAVVKDVEPMQEKAVSRLSDPELLGQYFNSGVVYLDLKKWADAKLTEKAL 201

Query: 177 DYIMQNRGRLILPDQDILNGLYANLVKPIPDEIYNYDARYSLIYQLKSRNEWDLEWVINH 236
             +M        PDQD++N L   +   +P E   Y+  Y++  +LK +   + + +I
Sbjct: 202 SILMSKDNVYKYPDQDVMNVLLKGMTLFLPRE---YNTIYTIKSELKDKTHQNYKKLITE 258

Query: 237 -TVFLHFAGRDKPWKK 251
            T+ +H+ G  KPW K
Sbjct: 259 STLLIHYTGATKPWHK 274
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1464

A DNA sequence (GBSx1550) was identified in *S.agalactiae* <SEQ ID 4499> which encodes the amino acid sequence <SEQ ID 4500>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1406(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1465

A DNA sequence (GBSx1551) was identified in *S.agalactiae* <SEQ ID 4501> which encodes the amino acid sequence <SEQ ID 4502>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.72   Transmembrane    7 -  23 (   1 -  28)
   INTEGRAL    Likelihood = -4.30   Transmembrane  222 - 238 ( 216 - 238)
   INTEGRAL    Likelihood = -3.66   Transmembrane  151 - 167 ( 140 - 170)
   INTEGRAL    Likelihood = -3.50   Transmembrane   35 -  51 (  34 -  58)
   INTEGRAL    Likelihood = -3.35   Transmembrane   71 -  87 (  69 -  88)
   INTEGRAL    Likelihood = -3.29   Transmembrane  113 - 129 ( 113 - 132)
   INTEGRAL    Likelihood = -2.81   Transmembrane  170 - 186 ( 168 - 190)
   INTEGRAL    Likelihood = -2.71   Transmembrane  198 - 214 ( 197 - 217)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07774 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 84/242 (34%), Positives = 147/242 (60%), Gaps = 16/242 (6%)

Query: 1   MVGLGTVINVILIIVGGFVGLFLKNFLKESLQKSLMQAMGVAVLFISISGVLEKMMLVEK 60
           MV +GTV+N   I++   +GL +KN + E ++ +LMQA+G+A++ + +     KM L  +
Sbjct: 1   MVLIGTVVNGAAIVIAALIGLLVKN-IPERVKTTLMQAIGLAIVLLGV-----KMGLQTE 54

Query: 61  SHLISNHTNMMIITLALGTVLGELLSLDSYIDKFGNYLKQKTGSGNDIKFVEAFVTSTCT 120
             LI      +I +L +G V+GE+++L+  +D  G +++ K G   D     AFVT+T
Sbjct: 55  QFLI------VICSLVIGGVIGEMINLEKRLDHLGRWIESKVGGKKDGSIATAFVTTTLI 108

Query: 121 VCIGAMAVVGSIQDGIAADHSILFAKGMLDMIIIAIMTVSLGKGALFSALPVALLQGSLT 180
             +GAMAV+G++  G+  DHS+L  K +LD  +  + T +LG G LFSA+PV L QGS+
Sbjct: 109 YVVGAMAVLGALDSGLRGDHSVLLTKALLDGFLAILFTSTLGIGVLFSAIPVVLYQGSIA 168

Query: 181 IVAF----FMGSLLNPSSLDYLNLVGNMLIFCVGVNLLFNLNIKVINMLPAIILAILWGS 236
           + A     ++ + L  S +  ++  G ++I  +G+NLL  +NI+V N+LP++++  +  +
Sbjct: 169 LFASQIDQYVPTALMDSFITEMSATGGVMIVAIGLNLLNVVNIRVANLLPSLVIVAVLVT 228

Query: 237 FI 238
           F+
Sbjct: 229 FV 230
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1466

A DNA sequence (GBSx1552) was identified in *S.agalactiae* <SEQ ID 4503> which encodes the amino acid sequence <SEQ ID 4504>. This protein is predicted to be alanyl-tRNA synthetase (alaS). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.41   Transmembrane  805 - 821 ( 804 - 822)

----- Final Results -----
               bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04986 GB:AP001511 alanyl-tRNA synthetase [Bacillus halodurans]
 Identities = 482/885 (54%), Positives = 618/885 (69%), Gaps = 27/885 (3%)

Query: 1   MKELSSAQIRQMWLDFWKSKGHSVEPSANLVPVNDPTLLWINSGVATLKKYFDGSVIPEN 60
           MK L+SAQ+RQM+LDF+K KGH VEPSA+LVP +DP+LLWINSGVATLKKYFDG VIPEN
Sbjct: 1   MKYLTSAQVRQMFLDFFKEKGHDVEPSASLVPHDDPSLLWINSGVATLKKYFDGRVIPEN 60

Query: 61  PRITNAQKSIRTNDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSPEWFDF 120
           PRITNAQKSIRTNDIENVGKTARHHT FEMLGNFSIGDYF++EAIEW +E LTS +W  F
Sbjct: 61  PRITNAQKSIRTNDIENVGKTARHHTFFEMLGNFSIGDYFKEEAIEWAWEFLTSEKWIGF 120

Query: 121 PKDKLYMTYYPDDKDSYNRWIA-CGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDF 179
```

```
            K+KL +T +P+D ++Y+ W    G+    ++ +E NFW+IG GPSGP+TEIF+DRG ++
Sbjct: 121 DKEKLSVTVHPEDDEAYSYWKEKIGIPEERIIRLEGNFWDIGEGPSGPNTEIFYDRGPEY 180

Query: 180 -----DPENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGL 234
                DPE      L    ENDRY+E+WN+V SQFN +P      Y  LP KNIDTG GL
Sbjct: 181 GDQPNDPE------LYPGGENDRYLEVWNLVFSQFNHNPD---GSYTPLPKKNIDTGMGL 231

Query: 235 ERLAAVMQGAKTNFETDLFMPIIREVEKLSGKTYDPDGD-NMSFKVIADHIRALSFAIGD 293
           ER+ +V+Q   TNFETDLFMPIIR  EK+SG  Y    + ++SFKVIADHIR ++FAIGD
Sbjct: 232 ERMVSVIQNVPTNFETDLFMPIIRATEKISGTEYGSHHEADVSFKVIADHIRTVTFAIGD 291

Query: 294 GALPGNEGRGYVLRRLLRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRDFIE 353
           GALP NEGRGYVLRRLLRRAV + +++GI+  F+Y+LVP VG IM  +YPEV EK  FI+
Sbjct: 292 GALPSNEGRGYVLRRLLRRAVRYAKQIGIDRPFMYELVPVVGDIMVDFYPEVKEKAAFIQ 351

Query: 354 KIVKREEETFARTIDAGSGHLDSLLAQLKAEGKDTLEGKDIFKLYDTYGFPVELTEELAE 413
           K+VK EEE F  T++ G   L+ ++ + K+EG  T+ G D+F+LYDTYGFPV+LTEE  E
Sbjct: 352 KVVKTEEERFHETLNEGLSILEKVIDKAKSEGASTISGSDVFRLYDTYGFPVDLTEEYVE 411

Query: 414 DAGYKIDHEGFKSAMKEQQDRARAAVVKGGSMGMQNETLAGIVEESRF-EYDTYSLESSL 472
           + G ++D +GF++ M+ Q++RAR A  + GSM +Q+E L  I  +S F  Y   S E+++
Sbjct: 412 EQGLQVDLDGFEAEMERQRERARTARQQAGSMQVQDEVLGQITVDSTFIGYKQLSTETTI 471

Query: 473 SVIIADNERTEAVSEGQ-ALLVFAQTPFYAEMGGQVADHGVIKNDKGDTVAEVVDVQKAP 531
             I+ D    + V  GQ A ++  +TPFYAE GGQVAD G+I+   G  V  V DVQKAP
Sbjct: 472 ETIVLDKTVADYVGAGQEAKVILKETPFYAESGGQVADKGIIRGANGFAV--VSDVQKAP 529

Query: 532 NGQPLHTVNVL-ASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNVIGEHATQAGS 590
           NGQ LHTV V   +L V       + +  R  + KNHTATHLLH AL +V+GEH  QAGS
Sbjct: 530 NGQHLHTVIVKEGTLQVNDQVQAIVEETERSGIVKNHTATHLLHRALKDVLGEHVNQAGS 589

Query: 591 LNEEEFLRFDFTHFEAVSNEELRHIEQEVNEQIWNDLTITTTETDVETAKEMGAMALFGE 650
           L  EE LRFDF+HF  V++EE   IE+ VNE+IW  + +  +   ++ AK +GAMALFGE
Sbjct: 590 LVSEERLRFDFSHFGQVTDEEKEKIERIVNEKIWQAIKVNISTKTLDEAKAIGAMALFGE 649

Query: 651 KYGKVVRVVQIGNYSVELCGGTHLNNSSEIGLFKIVKEEGIGSGTRRIIAVTGRQAFEAY 710
           KYG +VRVV++G+YS+ELCGG H+ N+SEIGLFKIV E GIG+G RRI AVTG++AF
Sbjct: 650 KYGDIVRVVEVGDYSIELCGGCHVTNTSEIGLFKIVSESGIGAGVRRIEAVTGKEAFLFM 709

Query: 711 RNQEDALKEIAATVKAPQLKDAAAKVQALSDSLRDLQKENVELKEKAAAAAAGDVFKDIQ 770
             Q D LKE AATVKA  +KD   +V+AL   +R+LQ+EN  L  K     AG +  ++Q
Sbjct: 710 AKQLDLLKETAATVKAKNVKDVPVRVEALQQQIRELQRENESLNAKLGNMEAGSLVNEVQ 769

Query: 771 EAKGVRFIASQVDVADAGALRTFADNWKQKDYSDVLVLVAAIGEKVNVLVASKTKDV--- 827
           + +GV  +A  +  AD   LR+  D  KQ+  S V+VL  A   KVN+ VA  TKD+
Sbjct: 770 KIEGVPVLAKAISGADMDGLRSIVDKLKQEIPSVVIVLGTASEGKVNI-VAGVTKDLINK 828

Query: 828 --HAGNMIKGLAPIVAGRGGGKPDMAMAGGSDASKIAELLAAVAE 870
             HAG ++K +A    G GGG+PDMA AGG    K+ + L+ V E
Sbjct: 829 GYHAGKLVKEVATRCGGGGGGRPDMAQAGGKQPEKLQDALSFVYE 873
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4505> which encodes the amino acid sequence <SEQ ID 4506>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.41    Transmembrane  805 - 821 ( 804 - 822)

----- Final Results -----
               bacterial membrane --- Certainty=0.2763(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 862/870 (99%), Positives = 864/870 (99%)
```

```
Query: 1   MKELSSAQIRQMWLDFWKSKGHSVEPSANLVPVNDPTLLWINSGVATLKKYFDGSVIPEN 60
           MKELSSAQIRQMWLDFWKSKGH VEPSANLVPVNDPTLLWINSGVATLKKYFDGSVIPEN
Sbjct: 1   MKELSSAQIRQMWLDFWKSKGHCVEPSANLVPVNDPTLLWINSGVATLKKYFDGSVIPEN 60

Query: 61  PRITNAQKSIRTNDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSPEWFDF 120
           PRITNAQKSIRTNDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSP+WFDF
Sbjct: 61  PRITNAQKSIRTNDIENVGKTARHHTMFEMLGNFSIGDYFRDEAIEWGFELLTSPDWFDF 120

Query: 121 PKDKLYMTYYPDDKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD 180
           PKDKLYMTYYPDDKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD
Sbjct: 121 PKDKLYMTYYPDDKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD 180

Query: 181 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAV 240
           PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAV
Sbjct: 181 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLAAV 240

Query: 241 MQGAKTNFETDLFMPIIREVEKLSGKTYDPDGDNMSFKVIADHIRALSFAIGDGALPGNE 300
           MQGAKTNFETDLFMPIIREVEKLSGKTYDPDGDNMSFKVIADHIRALSFAIGDGALPGNE
Sbjct: 241 MQGAKTNFETDLFMPIIREVEKLSGKTYDPDGDNMSFKVIADHIRALSFAIGDGALPGNE 300

Query: 301 GRGYVLRRLLRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRDFIEKIVKREE 360
           GRGYVLRRLLRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRDFIEKIVKREE
Sbjct: 301 GRGYVLRRLLRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRDFIEKIVKREE 360

Query: 361 ETFARTIDAGSGHLDSLLAQLKAEGKDTLEGKDIFKLYDTYGFPVELTEELAEDAGYKID 420
           ETFARTIDAGSGHLDSLLAQLKAEGKDTLEGKDIFKLYDTYGFPVELTEELAEDAGYKID
Sbjct: 361 ETFARTIDAGSGHLDSLLAQLKAEGKDTLEGKDIFKLYDTYGFPVELTEELAEDAGYKID 420

Query: 421 HEGFKSAMKEQQDRARAAVVKGGSMGMQNETLAGIVEESRFEYDTYSLESSLSVIIADNE 480
           HEGFKSAMKEQQDRARAAVVKGGSMGMQNETLAGIVEESRFEYDTYSLESSLSVIIADNE
Sbjct: 421 HEGFKSAMKEQQDRARAAVVKGGSMGMQNETLAGIVEESRFEYDTYSLESSLSVIIADNE 480

Query: 481 RTEAVSEGQALLVFAQTPFYAEMGGQVADHGVIKNDKGDTVAEVVDVQKAPNGQPLHTVN 540
           RTEAVSEGQALLVFAQTPFYAEMGGQVAD G IKNDKGDTVAEVVDVQKAPNGQPLHTVN
Sbjct: 481 RTEAVSEGQALLVFAQTPFYAEMGGQVADTGRIKNDKGDTVAEVVDVQKAPNGQPLHTVN 540

Query: 541 VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNVIGEHATQAGSLNEEEFLRFD 600
           VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNVIGEHATQAGSLNEEEFLRFD
Sbjct: 541 VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNVIGEHATQAGSLNEEEFLRFD 600

Query: 601 FTHFEAVSNEELRHIEQEVNEQIWNDLTITTTETDVETAKEMGAMALFGEKYGKVVRVVQ 660
           FTHFEAVSNEELRHIEQEVNEQIWN LTITTTETDVETAKEMGAMALFGEKYGKVVRVVQ
Sbjct: 601 FTHFEAVSNEELRHIEQEVNEQIWNALTITTTETDVETAKEMGAMALFGEKYGKVVRVVQ 660

Query: 661 IGNYSVELCGGTHLNNSSEIGLFKIVKEEGIGSGTRRIIAVTGRQAFEAYRNQEDALKEI 720
           IGNYSVELCGGTHLNNSSEIGLFKIVKEEGIGSGTRRIIAVTGRQAFEAYRNQEDALKEI
Sbjct: 661 IGNYSVELCGGTHLNNSSEIGLFKIVKEEGIGSGTRRIIAVTGRQAFEAYRNQEDALKEI 720

Query: 721 AATVKAPQLKDAAAKVQALSDSLRDLQKENVELKEKAAAAAAGDVFKDIQEAKGVRFIAS 780
           AATVKAPQLKDAAAKVQALSDSLRDLQKEN ELKEKAAAAAAGDVFKD+QEAKGVRFIAS
Sbjct: 721 AATVKAPQLKDAAAKVQALSDSLRDLQKENAELKEKAAAAAAGDVFKDVQEAKGVRFIAS 780

Query: 781 QVDVADAGALRTFADNWKQKDYSDVLVLVAAIGEKVNVLVASKTKDVHAGNMIKGLAPIV 840
           QVDVADAGALRTFADNWKQKDYSDVLVLVAAIGEKVNVLVASKTKDVHAGNMIK LAPIV
Sbjct: 781 QVDVADAGALRTFADNWKQKDYSDVLVLVAAIGEKVNVLVASKTKDVHAGNMIKELAPIV 840

Query: 841 AGRGGGKPDMAMAGGSDASKIAELLAAVAE 870
           AGRGGGKPDMAMAGGSDASKIAELLAAVAE
Sbjct: 841 AGRGGGKPDMAMAGGSDASKIAELLAAVAE 870
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1467

A DNA sequence (GBSx1553) was identified in *S.agalactiae* <SEQ ID 4507> which encodes the amino acid sequence <SEQ ID 4508>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2974(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9747> which encodes amino acid sequence <SEQ ID 9748> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15920 GB:Z99123 yxjI [Bacillus subtilis]
 Identities = 42/144 (29%), Positives = 73/144 (50%), Gaps = 2/144 (1%)

Query: 17  IKEKMFSLGGKFTITDLTGLPCYHVEGSLFPLPKTFKVFDEEEHLISQIEKKVLSFLPKF 76
           +K+KMFS    F I D      + VEG  F L  + ++ D     +  IE+K++S LP++
Sbjct: 6   MKQKMFSFKDAFHIYDRDEQETFKVEGRFFSLGDSLQMTDSSGKTLVSIEQKLMSLLPRY 65

Query: 77  NVTLANGNHFTIKKDFSFLKPHYTIEDLDMEVKGNFWDMDFQLLKDNQVIANISQQWFRM 136
            +++       + K  +F KP + I  L+ E+ G+ W  +FQL     V  ++S++W
Sbjct: 66  EISIGGKTVCEVTKKVTFSKPKFVISGLNWEIDGDLWRDEFQLTDGENVRMSVSKKWLSW 125

Query: 137 TSTYQVEVYSETYNDLTISLVIAI 160
             +Y +++  E   D+ I   IAI
Sbjct: 126 GDSYHLQIAYE--EDVLICTAIAI 147
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1468

A DNA sequence (GBSx1554) was identified in *S.agalactiae* <SEQ ID 4509> which encodes the amino acid sequence <SEQ ID 4510>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA36674 GB:AB016282 ORF17 [bacteriophage phi-105]
 Identities = 45/133 (33%), Positives = 74/133 (54%), Gaps = 5/133 (3%)

Query: 2   KYTYLALFEVDKENGGYNISFPDFHGAFSEADSLNEAIFNAREVLEIYTIMFEDEGKEFP 61
           +Y Y ALF+ D +  G  ++FPD  G  +  +S  EA+  A+E + ++   FE +G   P
Sbjct: 5   RYIYPALFDYDDD--GITVTFPDLPGCITFGNSGGEALTMAKEAMALHLYGFEQDGDIIP 62

Query: 62  KASSFKALASNLASDEDVIQAISVDTELVRERERSKIVNKTVTLPSWLVEVGKENKVNFS 121
           +A+  K +    A +   +  I       R    +  V KT+T+P W+ ++ KE+KVN+S
Sbjct: 63  EATPSKEIK---AEESQSVVLIETWMPPFRHDMENAAVKKTLTIPRWMDDIAKEHKVNYS 119
```

```
Query: 122 QLLQKAIREELQV 134
           QLLQ+AI+E L +
Sbjct: 120 QLLQEAIKEHLGI 132
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1469

A DNA sequence (GBSx1555) was identified in *S.agalactiae* <SEQ ID 4511> which encodes the amino acid sequence <SEQ ID 4512>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA25696 GB:AB010712 NADH oxidase/alkyl hydroperoxidase
           reductase [Streptococcus mutans]
 Identities = 383/509 (75%), Positives = 441/509 (86%)

Query: 1   MVLDKEIKAQLAQYLDLLESDIVLQADLGDNDNSQKVKDFLDEIVAMSDRISLESTHLKR 60
           M LD EIK QL QYL LLES+IVLQA L D+ NSQKVK+FL EIVAMS  ISLE   L R
Sbjct: 1   MALDAEIKEQLGQYLQLLESEIVLQAQLKDDANSQKVKEFLQEIVAMSPMISLEEKELPR 60

Query: 61  QPSFGIAKKGHESRVIFSGLPMGHEFTSFILALLQVSGRAPKVDEDIIKRIKGIEKTINL 120
            PSF IAKKG ES V F+GLP+GHEFTSFILALLQVSGR PKV+ DI+KRI+ +++ ++
Sbjct: 61  TPSFRIAKKGQESGVEFAGLPLGHEFTSFILALLQVSGRPPKVETDIVKRIQAVDEPMHF 120

Query: 121 ETYVSLTCHNCPDVVQAFNIMAVLNPNITHTMIEGGMYQDEVKSKGIMSVPTVYKDQEEF 180
           ETYVSLTCHNCPDVVQAFNIM+V+NPNI+HTM+EGGM++DE+++KGIMSVPTVYKD  EF
Sbjct: 121 ETYVSLTCHNCPDVVQAFNIMSVVNPNISHTMVEGGMFKDEIEAKGIMSVPTVYKDGTEF 180

Query: 181 TSGRATIEQLLEQLDGPLDAEAFADKGVYDVLVIGGGPAGNSAAIYAARKGLKTGILAET 240
           TSGRA+IEQLL+ + GPL  +AF DKGV+DVLVIGGGPAGNSAAIYAARKG+KTG+LAET
Sbjct: 181 TSGRASIEQLLDLIAGPLKEDAFDDKGVFDVLVIGGGPAGNSAAIYAARKGVKTGLLAET 240

Query: 241 FGGQVIETVGIENMIGTLYTEGPKLMAQIEEHTKSYDIDIIKSQLATGIEKKELVEVTLA 300
            GGQV+ETVGIENMIGT Y EGP+LMAQ+EEHTKSY +DI+K+  A  I+K +LVEV L
Sbjct: 241 MGGQVMETVGIENMIGTPYVEGPQLMAQVEEHTKSYSVDIMKAPRAKSIQKTDLVEVELD 300

Query: 301 NGAILQAKTAILALGAKWRNINVPGEEEFRNKGVTYCPHCDGPLFEGKDVAVIGGGNSGM 360
           NGA L+AKTA+LALGAKWR INVPGE+EF NKGVTYCPHCDGPLF  K VAVIGGGNSG+
Sbjct: 301 NGAHLKAKTAVLALGAKWRKINVPGEKEFFNKGVTYCPHCDGPLFTDKKVAVIGGGNSGL 360

Query: 361 EAALDLAGVTKHVTVLEFLPELKADQVLQERAAKTDNLTILKNVATKDIVGEDHVTGLNY 420
           EAA+DLAG+  HV +LEFLPELKAD++LQ+RA   DN+TIL NVATK+I+G DHV GL Y
Sbjct: 361 EAAIDLAGLASHVYILEFLPELKADKILQDRAEALDNITILTNVATKEIIGNDHVEGLRY 420

Query: 421 TDRDTNEEKHIDLEGVFVQIGLVPSTSWLKDSGIELNERQEIVVDKFGSTNIPGIFAAGD 480
           +DR TNEE  +DLEGVFVQIGLVPST WLKDSG+ LNE+ EI+V K G+TNIP IFAAGD
Sbjct: 421 SDRTTNEEYLLDLEGVFVQIGLVPSTDWLKDSGLALNEKGEIIVAKDGATNIPAIFAAGD 480

Query: 481 CTDAAYKQIIISMGSGATAAIGAFDYLIR 509
           CTD+AYKQIIISMGSGATAA+GAFDYLIR
Sbjct: 481 CTDSAYKQIIISMGSGATAALGAFDYLIR 509
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4513> which encodes the amino acid sequence <SEQ ID 4514>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0654(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 419/510 (82%), Positives = 472/510 (92%)

Query: 1   MVLDKEIKAQLAQYLDLLESDIVLQADLGDNDNSQKVKDFLDEIVAMSDRISLESTHLKR 60
           M L  +IK QLAQYL LLE+D+VLQ  LGDN+ SQKVKDF++EI AMS+RIS+E+  L R
Sbjct: 1   MALSPDIKEQLAQYLTLLEADLVLQVSLGDNEQSQKVKDFVEEIAAMSERISIENITLDR 60

Query: 61  QPSFGIAKKGHESRVIFSGLPMGHEFTSFILALLQVSGRAPKVDEDIIKRIKGIEKTINL 120
           QPSF +AKKGH S V+F+GLP+GHE TSFILALLQVSGRAPKVD+D+I RIK I++ ++
Sbjct: 61  QPSFKVAKKGHGSGVVFAGLPLGHELTSFILALLQVSGRAPKVDQDVIDRIKAIDRPLHF 120

Query: 121 ETYVSLTCHNCPDVVQAFNIMAVLNPNITHTMIEGGMYQDEVKSKGIMSVPTVYKDQEEF 180
           ETYVSLTCHNCPDVVQA NIM+VLN  I+HTM+EGGM+QDEVK+KGIMSVPTV+ D EEF
Sbjct: 121 ETYVSLTCHNCPDVVQALNIMSVLNDKISHTMVEGGMFQDEVKAKGIMSVPTVFLDGEEF 180

Query: 181 TSGRATIEQLLEQLDGPLDAEAFADKGVYDVLVIGGGPAGNSAAIYAARKGLKTGILAET 240
           TSGRATIEQLLEQ+ GPL  EAFADKG+YDVLVIGGGPAGNSAAIYAARKGLKTG+LAET
Sbjct: 181 TSGRATIEQLLEQIAGPLSEEAFADKGLYDVLVIGGGPAGNSAAIYAARKGLKTGLLAET 240

Query: 241 FGGQVIETVGIENMIGTLYTEGPKLMAQIEEHTKSYDIDIIKSQLATGIEKKELVEVTLA 300
           FGGQV+ETVGIENMIGTLYTEGPKLMA++E HTKSYD+DIIK+QLAT IEKKE +EVTLA
Sbjct: 241 FGGQVMETVGIENMIGTLYTEGPKLMAEVEAHTKSYDVDIIKAQLATSIEKKENIEVTLA 300

Query: 301 NGAILQAKTAILALGAKWRNINVPGEEEFRNKGVTYCPHCDGPLFEGKDVAVIGGGNSGM 360
           NGA+LQAKTAILALGAKWRNINVPGE+EFRNKGVTYCPHCDGPLFEGKDVAVIGGGNSG+
Sbjct: 301 NGAVLQAKTAILALGAKWRNINVPGEDEFRNKGVTYCPHCDGPLFEGKDVAVIGGGNSGL 360

Query: 361 EAALDLAGVTKHVTVLEFLPELKADQVLQERAAKTDNLTILKNVATKDIVGEDHVTGLNY 420
           EAALDLAG+ KHV VLEFLPELKAD+VLQ+RAAKT+N+TI+KNVATKDIVGEDHVTGLNY
Sbjct: 361 EAALDLAGLAKHVYVLEFLPELKADKVLQDRAAKTNNMTIIKNVATKDIVGEDHVTGLNY 420

Query: 421 TDRDTNEEKHIDLEGVFVQIGLVPSTSWLKDSGIELNERQEIVVDKFGSTNIPGIFAAGD 480
           T+RD+ E+KH+DLEGVFVQIGLVP+T+WLKDSG+ L +R EI+VDK GSTNIPGIFAAGD
Sbjct: 421 TERDSGEDKHLDLEGVFVQIGLVPNTAWLKDSGVNLTDRGEIIVDKHGSTNIPGIFAAGD 480

Query: 481 CTDAAYKQIIISMGSGATAAIGAFDYLIRQ 510
           CTD+AYKQIIISMGSGATAAIGAFDYLIRQ
Sbjct: 481 CTDSAYKQIIISMGSGATAAIGAFDYLIRQ 510
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1470

A DNA sequence (GBSx1556) was identified in *S.agalactiae* <SEQ ID 4515> which encodes the amino acid sequence <SEQ ID 4516>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2906(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA25695 GB:AB010712 alkyl hydroperoxidase [Streptococcus mutans]
 Identities = 167/186 (89%), Positives = 179/186 (95%)

Query: 1   MSLVGKEIIEFSAQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTELGDLQEQYET 60
            MSLVGKE++EFSAQAYH G+F+TV NEDVKGKWAVFCFYPADFSFVCPTELGDLQEQY T
Sbjct: 1   MSLVGKEMVEFSAQAYHQGEFVTVNNEDVKGKWAVFCFYPADFSFVCPTELGDLQEQYAT 60

Query: 61  LKSLDVEVYSVSTDTHFVHKAWHDDSDVVGTITYPMIGDPSHLISQGFDVLGQDGLAQRG 120
            L+SL VEVYSVSTDTHFVHKAWHDDSDVVGTITY MIGDPSH++SQGF+VLG+DGLAQRG
Sbjct: 61  LQSLGVEVYSVSTDTHFVHKAWHDDSDVVGTITYTMIGDPSHVLSQGFEVLGEDGLAQRG 120

Query: 121 TFIIDPDGVIQMMEINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKEGAETLTPSL 180
            TFI+DPDG+IQMME+NADGIGRDASTLIDKVRAAQYIRQH GEVCPAKWKEGAETL PSL
Sbjct: 121 TFIVDPDGIIQMMEVNADGIGRDASTLIDKVRAAQYIRQHPGEVCPAKWKEGAETLKPSL 180

Query: 181 DLVGKI 186
            DLVGKI
Sbjct: 181 DLVGKI 186
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4517> which encodes the amino acid sequence <SEQ ID 4518>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3022(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 173/186 (93%), Positives = 181/186 (97%)

Query: 1   MSLVGKEIIEFSAQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTELGDLQEQYET 60
            MSL+GKEI EFSAQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTELGDLQEQYET
Sbjct: 1   MSLIGKEIAEFSAQAYHDGKFITVTNEDVKGKWAVFCFYPADFSFVCPTELGDLQEQYET 60

Query: 61  LKSLDVEVYSVSTDTHFVHKAWHDDSDVVGTITYPMIGDPSHLISQGFDVLGQDGLAQRG 120
            LKSL VEVYSVSTDTHFVHKAWHDDSDVVGTITYPMIGDPSHLISQ F+VLG+DGLAQRG
Sbjct: 61  LKSLGVEVYSVSTDTHFVHKAWHDDSDVVGTITYPMIGDPSHLISQAFEVLGEDGLAQRG 120

Query: 121 TFIIDPDGVIQMMEINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKEGAETLTPSL 180
            TFI+DPDG+IQMMEINADGIGRDASTLIDK+ AAQY+R+H GEVCPAKWKEGAETLTPSL
Sbjct: 121 TFIVDPDGIIQMMEINADGIGRDASTLIDKIHAAQYVRKHPGEVCPAKWKEGAETLTPSL 180

Query: 181 DLVGKI 186
            DLVGKI
Sbjct: 181 DLVGKI 186
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1471

A DNA sequence (GBSx1557) was identified in *S.agalactiae* <SEQ ID 4519> which encodes the amino acid sequence <SEQ ID 4520>. This protein is predicted to be 30S ribosomal protein S2 (rpsB). Analysis of this protein sequence reveals the following:

```
Possible site: 60
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA50276 GB:X70925 30S ribosomal protein [Pediococcus
           acidilactici]
 Identities = 190/260 (73%), Positives = 226/260 (86%), Gaps = 4/260 (1%)

Query: 1   MAVISMKQLLEAGVHFGHQTRRWNPKMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA 60
           M+VISMKQLLEAGVHFGHQTRRWNPKM  +IFTERNGI++IDLQ+TVKL D AY FV+D
Sbjct: 1   MSVISMKQLLEAGVHFGHQTRRWNPKMKPFIFTERNGIYIIDLQKTVKLIDNAYNFVKDV 60

Query: 61  AANDAVILFVGTKKQAAEAVAEEAKRAGQYFINHRWLGGTLTNWGTIQKRIARLKEIKRM 120
           AAND V+LFVGTKKQA  A+ EEAKRAGQ+++NHRWLGGTLTNW TIQKRI RLK++K+M
Sbjct: 61  AANDGVVLFVGTKKQAQTAIEEEAKRAGQFYVNHRWLGGTLTNWNTIQKRIKRLKDLKKM 120

Query: 121 EEEGTFELLPKKEVALLNKQRARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG 180
           EE+GTF+ LPKKEVALLNKQ+ +LEKFLGGIEDMP IPDV++VVDP KEQIA+KEA+KL
Sbjct: 121 EEDGTFDRLPKKEVALLNKQKDKLEKFLGGIEDMPHIPDVLFVVDPRKEQIAIKEAQKLN 180

Query: 181 IPVVAMVDTNADPDDIDVIIPANDDAIRAVKLITSKLADAVIEGRQGEDADV----DFAQ 236
           IPVVAMVDTN DPD +DVIIP+NDDAIRAV+LITSK+ADAV+EGRQGED +     + A+
Sbjct: 181 IPVVAMVDTNTDPDQVDVIIPSNDDAIRAVRLITSKMADAVVEGRQGEDDEAVQQEEVAE 240

Query: 237 EAQADSIEEIVEVVEGSNND 256
               DS+E++ + VE  +N+
Sbjct: 241 GVSKDSLEDLKKTVEEGSNE 260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4521> which encodes the amino acid sequence <SEQ ID 4522>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 241/254 (94%), Positives = 248/254 (96%)

Query: 1   MAVISMKQLLEAGVHFGHQTRRWNPKMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA 60
           MAVISMKQLLEAGVHFGHQTRRWNPKMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA
Sbjct: 1   MAVISMKQLLEAGVHFGHQTRRWNPKMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA 60

Query: 61  AANDAVILFVGTKKQAAEAVAEEAKRAGQYFINHRWLGGTLTNWGTIQKRIARLKEIKRM 120
           AANDAVILFVGTKKQAAEAVA+EA RAGQYFINHRWLGGTLTNWGTIQKRIARLKEIKRM
Sbjct: 61  AANDAVILFVGTKKQAAEAVADEATRAGQYFINHRWLGGTLTNWGTIQKRIARLKEIKRM 120

Query: 121 EEEGTFELLPKKEVALLNKQRARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG 180
           EEEGTF++LPKKEVALLNKQRARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG
Sbjct: 121 EEEGTFDVLPKKEVALLNKQRARLEKFLGGIEDMPRIPDVMYVVDPHKEQIAVKEAKKLG 180

Query: 181 IPVVAMVDTNADPDDIDVIIPANDDAIRAVKLITSKLADAVIEGRQGEDADVDFAQEAQA 240
           IPVVAMVDTNADPDDID+IIPANDDAIRAVKLIT+KLADA+IEGRQGEDADV F  + QA
Sbjct: 181 IPVVAMVDTNADPDDIDIIIPANDDAIRAVKLITAKLADAIIEGRQGEDADVAFEADTQA 240

Query: 241 DSIEEIVEVVEGSN 254
           DSIEEIVEVVEG N
Sbjct: 241 DSIEEIVEVVEGDN 254
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1472

A DNA sequence (GBSx1558) was identified in *S.agalactiae* <SEQ ID 4523> which encodes the amino acid sequence <SEQ ID 4524>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2648(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB73435 GB:AL139077 elongation factor TS [Campylobacter jejuni]
 Identities = 169/358 (47%), Positives = 226/358 (62%), Gaps = 19/358 (5%)

Query: 1   MAEITAKLVKELREKSGAGVMDAKKALVETDGDLDKAIELLREKGMAKAAKKADRVAAEG 60
           M EITA +VKELRE +GAG+MD K AL ET+GD DKA++LLREKG+ KAAKKADR+AAEG
Sbjct: 1   MTEITAAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKAAKKADRLAAEG 60

Query: 61  LTGVYV--DGNVAAVIEVNAETDFVAKNDQFVTLVNETAKVIAEGRPSNNEEALALTMPS 118
           L  V V  D   A V E+N+ETDFVAKNDQF+ L  +T   I      + EE  + T+ +
Sbjct: 61  LVSVKVSDDFTSATVSEINSETDFVAKNDQFIALTKDTTAHIQSNSLQSVEELHSSTI-N 119

Query: 119 GETLEQAFVTATATIGEKISFRRFALVEKTDEQHFGAYQHNGGRIGVITV-------VEG 171
           G   E+   +  ATIGE +  RRFA ++         Y H  GR+GV+         V
Sbjct: 120 GVKFEEYLKSQIATIGENLVVRRFATLKAGANGVVNGYIHTNGRVGVVIAAACDSAEVAS 179

Query: 172 GDDALAKQVSMHVAAMKPTVLSYTELDAQFVHDELAQLNHKIEQDNESRAMV---NKPAL 228
               L +Q+ MH+AAM+P+ LSY +LD  FV +E   L  ++E++NE R  +   NKP
Sbjct: 180 KSRDLLRQICMHIAAMRPSYLSYEDLDMTFVENEYKALVAELEKENEERRRLKDPNKPEH 239

Query: 229 PFLKYGSKAQLTDEVIAQAEEDIKAELAAEGKPEKIWDKIVPGKMDRFMLDNTKVDQEYT 288
              ++ S+ QL+D ++ +AEE IK EL A+GKPEKIWD I+PGKM+ F+ DN+++D + T
Sbjct: 240 KIPQFASRKQLSDAILKEAEEKIKEELKAQGKPEKIWDNIIPGKMNSFIADNSQLDSKLT 299

Query: 289 LLAQVYIMDDSKTVEAYLESV------NAKAVAFVRFEVGEGIEKASNDFEAEVAATM 340
           L+ Q Y+MDD KTVE  +           K V F+ FEVGEG+EK + DF AEVAA +
Sbjct: 300 LMGQFYVMDDKKTVEQVIAEKEKEFGGKIKIVEFICFEVGEGLEKKTEDFAAEVAAQL 357
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4525> which encodes the amino acid sequence <SEQ ID 4526>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3942(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 307/344 (89%), Positives = 327/344 (94%)

Query: 1   MAEITAKLVKELREKSGAGVMDAKKALVETDGDLDKAIELLREKGMAKAAKKADRVAAEG 60
           MAEITAKLVKELREKSGAGVMDAKKALVETDGD+DKA+ELLREKGMAKAAKKADRVAAEG
Sbjct: 33  MAEITAKLVKELREKSGAGVMDAKKALVETDGDMDKAVELLREKGMAKAAKKADRVAAEG 92
```

```
Query: 61  LTGVYVDGNVAAVIEVNAETDFVAKNDQFVTLVNETAKVIAEGRPSNNEEALALTMPSGE 120
           LTGVYV GNVAAV+EVNAETDFVAKN QFV LVN TAKVIAEG+P+NN+EALAL MPSGE
Sbjct: 93  LTGVYVHGNVAAVVEVNAETDFVAKNAQFVELVNATAKVIAEGKPANNDEALALVMPSGE 152

Query: 121 TLEQAFVTATATIGEKISFRRFALVEKTDEQHFGAYQHNGGRIGVITVVEGGDDALAKQV 180
           TL +A+V ATATIGEKISFRRFAL+EK DEQHFGAYQHNGGRIGVI+VVEGGDDALAKQV
Sbjct: 153 TLAEAYVNATATIGEKISFRRFALIEKADEQHFGAYQHNGGRIGVISVVEGGDDALAKQV 212

Query: 181 SMHVAAMKPTVLSYTELDAQFVHDELAQLNHKIEQDNESRAMVNKPALPFLKYGSKAQLT 240
           SMH+AAMKPTVLSYTELDAQF+ DELAQLNH IE DNESRAMV+KPALPFLKYGSKAQL+
Sbjct: 213 SMHIAAMKPTVLSYTELDAQFIKDELAQLNHAIELDNESRAMVDKPALPFLKYGSKAQLS 272

Query: 241 DEVIAQAEEDIKAELAAEGKPEKIWDKIVPGKMDRFMLDNTKVDQEYTLLAQVYIMDDSK 300
           D+VI  AE DIKAELAAEGKPEKIWDKI+PGKMDRFMLDNTKVDQ YTLLAQVYIMDDSK
Sbjct: 273 DDVITAAEADIKAELAAEGKPEKIWDKIIPGKMDRFMLDNTKVDQAYTLLAQVYIMDDSK 332

Query: 301 TVEAYLESVNAKAVAFVRFEVGEGIEKASNDFEAEVAATMAAAL 344
           TVEAYL+SVNAKA+AF RFEVGEGIEK +NDFE+EVAATMAAAL
Sbjct: 333 TVEAYLDSVNAKAIAFARFEVGEGIEKKANDFESEVAATMAAAL 376
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1473

A DNA sequence (GBSx1559) was identified in *S.agalactiae* <SEQ ID 4527> which encodes the amino acid sequence <SEQ ID 4528>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1312(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1474

A DNA sequence (GBSx1560) was identified in *S.agalactiae* <SEQ ID 4529> which encodes the amino acid sequence <SEQ ID 4530>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.86   Transmembrane  128 - 144 ( 124 - 152)
   INTEGRAL    Likelihood = -4.57   Transmembrane   35 -  51 (  33 -  53)
   INTEGRAL    Likelihood = -4.04   Transmembrane   92 - 108 (  87 - 111)

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04953 GB:AP001511 small multidrug export related protein
          [Bacillus halodurans]
```

```
 Identities = 47/137 (34%), Positives = 71/137 (51%), Gaps = 5/137 (3%)

Query: 12  IPLVELRGAVPFAIANGIPLWEALAIGVVGNMLPVPIIFFFARKVLEWGADKPYTGKFFT 71
            +P+VELRG +P  +  G+  WEAL  G++GN+LP+  I    R +  W     +  + +
Sbjct: 1   MPIVELRGGIPLGVVLGLSPWEALLFGIIGNLLPIVPILLLFRPISGWMLRFKWYQRLYD 60

Query: 72  WCLKKGHSGGQKLEKVAGEKGLFIALLLFVGIPLPGTGAWTGTLAASLLDWEFKHSVIAV 131
            W   +       +EK        I L+LF  +PLP TGA++  LAA L    F+ +  AV
Sbjct: 61  WLYNRTMKKSNNVEKFGA-----IGLILFTAVPLPTTGAYSACLAAVLFFIPFRFAFFAV 115

Query: 132 MLGVILAGCIMGTLSII 148
              GV++AG +M   S I
Sbjct: 116 SAGVVIAGIVMTLFSYI 132
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8817> and protein <SEQ ID 8818> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      3.98
GvH: Signal Score (-7.5): -2.35
     Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 3 value:  -7.86 threshold:  0.0
   INTEGRAL    Likelihood = -7.86   Transmembrane  128 - 144 ( 124 - 152)
   INTEGRAL    Likelihood = -4.57   Transmembrane   35 -  51 (  33 -  53)
   INTEGRAL    Likelihood = -4.04   Transmembrane   92 - 108 (  87 - 111)
   PERIPHERAL  Likelihood = 12.20     109
 modified ALOM score:   2.07

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 105-109
```

The protein has homology with the following sequences in the databases:

```
186       216       246       276       306       336       366       396
LTIIISNF*KIRK*NLSKDSKTRMTADFSCHY*KDKIKWNNTIERFYLMNYIITFLISMIPLVELRGAVPFAIANGIPLW
                                                           :|: |||||:|:|:  |
                                                           MPFSELRGAIPLALYFGFSPA
                                                                   10        20

426       456       486       516       546       576       591       621
EALAIGVVGNMLPVPIIFFFARKVLEWGADKPYTGKFFTWCLKKGHSGGQKLEKVAGEKGL-----FIALLLFVGIPLPG
||  : |:||:|||| :::|   ::          : :           : :|:|   ||:     :: | :|  ||||
EAYLLSVLGNILPVPFLLLFLDYLVRIATKVELLARIYR----------RVVERVERRKGVVERYGYLGLTIFVAIPLPV
                 40        50        60                  70        80        90

651       681       711       741       771       801       831       861
TGAWTGTLAASLLDWEFKHSVIAVMLGVILAGCIMGTLSIIGFNLF*KS*GEMTVSPF*YLPIHQFDSKIRHLT*AKCLI
|||||||| | ||      : : :  || :|| ::   ||    |:
TGAWTGTLLAFLLQLNRLKAFLFISAGVCIAGVVVLLASIGIIRLL
                 110       120       130
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1475

A DNA sequence (GBSx1561) was identified in *S.agalactiae* <SEQ ID 4531> which encodes the amino acid sequence <SEQ ID 4532>. This protein is predicted to be CtsR protein (ctsR). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3672(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB91548 GB:AJ249133 CtsR protein [Lactococcus lactis]
 Identities = 74/146 (50%), Positives = 103/146 (69%), Gaps = 3/146 (2%)

Query: 4    KNTSDNIEEYIKSLLEQSGIAEIKRSNLADTFQVVPSQINYVIKTRFTESRGYVVESKRG 63
            KNTSD IE Y++ LLE++ + EIKR++LA+ F VVPSQINYVIKTRFT S+G+ VESKRG
Sbjct: 5    KNTSDIIEAYLRQLLEEAQVIEIKRADLANQFDVVPSQINYVIKTRFTASKGFDVESKRG 64

Query: 64   GGGYIRIAKVHFSDQHQLFGNMLSTIGERISEQVFDDLIQLLFDEEIITEREGNLILATS 123
            GGGYI+I K  +S +H+    +   +   +S +   D++QLLFDE+++TEREGNL+L
Sbjct: 65   GGGYIKIVKYQYSARHEFLTALYQKVPANLSSKAAHDIVQLLFDEKVLTEREGNLLLLVI 124

Query: 124 GDDVLGEQASVIRARMLRKLLQRLDR 149
             D   G  +   R  M++ ++ RLDR
Sbjct: 125 TD---GAISPFTRGIMMKSIINRLDR 147
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4533> which encodes the amino acid sequence <SEQ ID 4534>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2514(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 117/151 (77%), Positives = 131/151 (86%)

Query: 1    MAIKNTSDNIEEYIKSLLEQSGIAEIKRSNLADTFQVVPSQINYVIKTRFTESRGYVVES 60
            M  KNTSD+IEEYIK LL +SGIAEIKRS LAD+FQVVPSQINYVIKTRFTESRGY VES
Sbjct: 1    MPTKNTSDSIEEYIKELLAKSGIAEIKRSMLADSFQVVPSQINYVIKTRFTESRGYEVES 60

Query: 61   KRGGGGYIRIAKVHFSDQHQLFGNMLSTIGERISEQVFDDLIQLLFDEEIITEREGNLIL 120
            KRGGGGYIRIAKVHFSD+H L GN+++TI + ISEQVF D IQLLFDE ++TEREGN+IL
Sbjct: 61   KRGGGGYIRIAKVHFSDKHHLIGNLMATIEDCISEQVFTDSIQLLFDEHLLTEREGNIIL 120

Query: 121 ATSGDDVLGEQASVIRARMLRKLLQRLDRKG 151
            A + DDVLG   S IRARML +LLQR+DRKG
Sbjct: 121 AVASDDVLGTDGSTIRARMLYRLLQRIDRKG 151
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1476

A DNA sequence (GBSx1562) was identified in *S.agalactiae* <SEQ ID 4535> which encodes the amino acid sequence <SEQ ID 4536>. This protein is predicted to be ClpC (clpB-1). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.34    Transmembrane   32 -  48 (  32 -  49)

----- Final Results -----
               bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD01783 GB:AF023422 ClpC [Lactococcus lactis]
 Identities = 401/831 (48%), Positives = 571/831 (68%), Gaps = 52/831 (6%)

Query: 4   YSIKLQEVFRLAQFQAARYESHYLESWHLLLAMVLVHDSVAGLTFAEYE---SEVAIEEY 60
            Y+  L  +F  A   A +Y+   +ES HLL AM     S+A    A      S++ I+
Sbjct: 8   YTPTLDRIFEKAAEYAHQYQYGTIESAHLLAAMATTSGSIAYSILAGMNVDSSDLLIDLE 67

Query: 61  EAATILALGRAPKEEITNYQFLEQSPALKKILKLAENISIVVGAEDVGTEHVLLAMLVNK 120
            + ++ + + R+          L  SP  ++++ +A  +++   AE VGTEH+L A+L  +
Sbjct: 68  DLSSHVKVKRSE---------LRFSPRAEEVVTVASFLAVHNNAEAVGTEHLLYALLQVE 118

Query: 121 DLLATRILELVGFRGQDDGESVRMVDLRKALERHAGF-TKDDIKAIYELRNPKKAKSGAS 179
            D    ++L+L         + + +V LRK +E+  G    ++ KA+  +   K AK  A
Sbjct: 119 DGFGLQLLKL---------QKINIVSLRKEIEKRTGLIVPENKKAVTPMSKRKMAKGVAE 169

Query: 180 FSDMMKPPSTAGDLADFTRDLSQMAVDGEIEPVIGRDKEISRMVQVLSRKTKNNPVLVGD 239
                    S+   L   + DL++ A  G+++P+IGR+ E+ R++ +LSR+TKNNPVLVG+
Sbjct: 170 -------NSSTPTLDSVSSDLTEAARSGKLDPMIGREAEVDRLIHILSRRTKNNPVLVGE 222

Query: 240 AGVGKTALAYGLAQRIANGNIPYELRDMRVLELDMMSVVAGTRFRGDFEERMNQIIADIE 299
             GVGK+A+  GLAQRI NG +P  L + R++ L+M +VVAGT+FRG+FE+R+  I+ ++
Sbjct: 223 PGVGKSAIIEGLAQRIVNGQVPIGLMNSRIMALNMATVVAGTKFRGEFEDRLTAIVEEVS 282

Query: 300 EDGHIILFIDELHTIMGSGSGIDSTLDAANILKPALARGTLRTVGATTQEEYQKHIEKDA 359
             D  +I+FIDELHTI+G+G G+DS  DAANILKPALARG  + VGATT  EYQK+IEKD
Sbjct: 283 ADPDVIIFIDELHTIIGAGGGMDSVNDAANILKPALARGDFQMVGATTYHEYQKYIEKDE 342

Query: 360 ALSRRFAKVLVEEPNLEDAYEILLGLKPAYEAFHNVTISDEAVMTAVKVAHRYLTSKNLP 419
            AL RR A++ V+EP+ ++A  IL GL+  +E +H V  +D+A+ +AV ++ RY+TS+ LP
Sbjct: 343 ALERRLARINVDEPSPDEAIAILQGLREKFEDYHQVKFTDQAIKSAVTLSVRYMTSRKLP 402

Query: 420 DSAIDLLDEASATVQMMIKKNAPSLLT----------EVDQAILDDDMKSA--------- 460
            D AIDLLDEA+A V++++K     ++           E+ +A++  D+K++
Sbjct: 403 DKAIDLLDEAAARVKILLKTKKQNVFELEKDFVKAQEELAEAVIKLDVKASRIKEKAVEK 462

Query: 461 --SKALKASYKGKKRKPIAVTEDHIMATLSRLSGIPVEKLTQADSKKYLNLEKELHKRVI 518
               K  K S K +KR+   VT+  ++A  S L+G+P+ ++T+++S + +NLEKELHKRV+
Sbjct: 463 ISDKIYKFSIKEEKRQE--VTDQAVIAVASTLTGVPITQMTKSESDRLINLEKELHKRVV 520

Query: 519 GQDDAVTAISRAIRRNQSGIRTGKRPIGSFMFLGPTGVGKTELAKALAEVLFDDESALIR 578
            GQ++A++A+SRAIRR +SG+   +RP+GSFMFLGPTGVGKTELAKALA+ +F  E  +IR
Sbjct: 521 GQEEAISAVSRAIRRARSGVADSRRPMGSFMFLGPTGVGKTELAKALADSVFGSEDNMIR 580

Query: 579 FDMSEYMEKFAASHLNGAPPGYVGYDEGGELTEKVRNKPYSVLLFDEVEKAHPDIFNVLL 638
             DMSE+MEK + S L GAPPGYVGYDEGG+LTE+VRNKPYSV+L DEVEKAH D+FN++L
Sbjct: 581 VDMSEFMEKHSTSRLIGAPPGYVGYDEGGQLTERVRNKPYSVVLLDEVEKAHLDVFNIML 640

Query: 639 QVLDDGVLTDSRGRKVDFSNTIIIMTSNLGATALRDDKTVGFGAKDISHDYTAMQKRIME 698
            Q+LDDG +TD++GRKVDF NTIIIMTSNLGATALRDDKTVGFGAK+I+ DY+AMQ RI+E
Sbjct: 641 QILDDGFVTDTKGRKVDFRNTIIIMTSNLGATALRDDKTVGFGAKNITADYSAMQSRILE 700
```

```
Query: 699 ELKKAYRPEFINRIDEKVVFHSLSQDNMREVVKIMVKPLILALKDKGMDLKFQPSALKHL 758
           ELK+ YRPEF+NRIDE +VFHSL    + ++VKIM K LI  L ++ + +K  PSA+K +
Sbjct: 701 ELKRHYRPEFLNRIDENIVFHSLESQEIEQIVKIMSKSLIKRLAEQDIHVKLTPSAIKLI 760

Query: 759 AEDGYDIEMGARPLRRTIQTQVEDHLSELLLANQVKEGQVIKIGVSKGKLK 809
           AE G+D E GARPLR+ +Q +VED LSE LL+ ++K G  I IG S  K+K
Sbjct: 761 AEVGFDPEYGARPLRKALQKEVEDLLSEQLLSGEIKAGNHISIGASNKKIK 811
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4537> which encodes the amino acid sequence <SEQ ID 4538>. Analysis of this protein sequence reveals the following:

```
      Possible site: 44

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.75    Transmembrane    32 -  48 (  32 -  48)

----- Final Results -----
            bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 285-287
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 618/814 (75%), Positives = 716/814 (87%), Gaps = 1/814 (0%)

Query: 1   MSHYSIKLQEVFRLAQFQAARYESHYLESWHLLLAMVLVHDSVAGLTFAEYESEVAIEEY 60
           M  YS K+Q++FR AQFQAAR++SH LE+WH+LLAMV V +S+A +  +EY+++VAIEEY
Sbjct: 1   MIMYSTKMQDIFRQAQFQAARFDSHCLETWHVLLAMVAVDNSLANMILSEYDAQVAIEEY 60

Query: 61  EAATILALGRAPKEEITNYQFLEQSPALKKILKLAENISIVVGAEDVGTEHVLLAMLVNK 120
           EAA ILA+G+ PKE+++   F  QS  L  +L  A+ IS +   ++VG+EHVL A+L+N
Sbjct: 61  EAAAILAMGKTPKEQLSRVDFRPQSKTLTNLLAFAQAISQITRDQEVGSEHVLFAILLNP 120

Query: 121 DLLATRILELVGFRGQDDGESV-RMVDLRKALERHAGFTKDDIKAIYELRNPKKAKSGAS 179
           D++A+R+LE+ G++ +D+G    R+ DLRKA+ERHAG++K+ IKAI+ELR PKK K+  +
Sbjct: 121 DIMASRLLEIAGYQIKDNGNGQPRLADLRKAIERHAGYSKEMIKAIHELRKPKKTKTQGT 180

Query: 180 FSDMMKPPSTAGDLADFTRDLSQMAVDGEIEPVIGRDKEISRMVQVLSRKTKNNPVLVGD 239
           FSDMMKPPSTAG+L+DFTRDL++MA  G +E VIGRD+E+SRM+QVLSRKTKNNPVLVGD
Sbjct: 181 FSDMMKPPSTAGELSDFTRDLTEMARQGLLESVIGRDQEVSRMIQVLSRKTKNNPVLVGD 240

Query: 240 AGVGKTALAYGLAQRIANGNIPYELRDMRVLELDMMSVVAGTRFRGDFEERMNQIIADIE 299
           AGVGKTALAYGLAQRIANG IPYEL++MRVLELDMMSVVAGTRFRGDFEERMNQII DIE
Sbjct: 241 AGVGKTALAYGLAQRIANGAIPYELKEMRVLELDMMSVVAGTRFRGDFEERMNQIIDDIE 300

Query: 300 EDGHIILFIDELHTIMGSGSGIDSTLDAANILKPALARGTLRTVGATTQEEYQKHIEKDA 359
            DG IILF+DELHTIMGSGSGIDSTLDAANILKPAL+RGTL  VGATTQEEYQKHIEKDA
Sbjct: 301 ADGQIILFVDELHTIMGSGSGIDSTLDAANILKPALSRGTLHMVGATTQEEYQKHIEKDA 360

Query: 360 ALSRRFAKVLVEEPNLEDAYEILLGLKPAYEAFHNVTISDEAVMTAVKVAHRYLTSKNLP 419
           ALSRRFAK+L+EEPN EDAY+IL+GLK +YE +HNV+IS+EAV TAVK+AHRYLTSKNLP
Sbjct: 361 ALSRRFAKILIEEPNTEDAYQILMGLKLSYETYHNVSISNEAVKTAVKMAHRYLTSKNLP 420

Query: 420 DSAIDLLDEASATVQMMIKKNAPSLLTEVDQAILDDDMKSASKALKASYKGKKRKPIAVT 479
           DSAIDLLDEASA VQ M+KK+AP  LT +DQA+++ DMK  S+ L    KG+ RKP  VT
Sbjct: 421 DSAIDLLDEASAAVQNMVKKSAPETLTPIDQALINGDMKKVSRLLAKEAKGQMRKPTPVT 480

Query: 480 EDHIMATLSRLSGIPVEKLTQADSKKYLNLEKELHKRVIGQDDAVTAISRAIRRNQSGIR 539
           ED I+ATLS+LSGIP+EKLTQADSKKYLNLEKELHKRVIGQD AVTAISRAIRRNQSGIR
Sbjct: 481 EDDILATLSKLSGIPLEKLTQADSKKYLNLEKELHKRVIGQDAAVTAISRAIRRNQSGIR 540

Query: 540 TGKRPIGSFMFLGPTGVGKTELAKALAEVLFDDESALIRFDMSEYMEKFAASHLNGAPPG 599
           TGKRPIGSFMFLGPTGVGKTELAKALAEVLFDDE+ALIRFDMSEYMEKFAAS LNGAPPG
Sbjct: 541 TGKRPIGSFMFLGPTGVGKTELAKALAEVLFDDEAALIRFDMSEYMEKFAASRLNGAPPG 600
```

```
Query: 600 YVGYDEGGELTEKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDGVLTDSRGRKVDFSNT 659
           YVGYDEGGELT+KVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDG+LTDSRGRKVDFSNT
Sbjct: 601 YVGYDEGGELTQKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDGILTDSRGRKVDFSNT 660

Query: 660 IIIMTSNLGATALRDDKTVGFGAKDISHDYTAMQKRIMEELKKAYRPEFINRIDEKVVFH 719
           IIIMTSNLGATALRDDKTVGFG KDI  D+ AM+KRI+EEL+K YRPEFINRIDEKVVFH
Sbjct: 661 IIIMTSNLGATALRDDKTVGFGVKDIHQDHQAMEKRILEELRKTYRPEFINRIDEKVVFH 720

Query: 720 SLSQDNMREVVKIMVKPLILALKDKGMDLKFQPSALKHLAEDGYDIEMGARPLRRTIQTQ 779
           SL+QDNMR+VVKIMV+PLI  L +KG+ LK QP ALKHL+E GYD  MGARPLRRT+QT+
Sbjct: 721 SLTQDNMRDVVKIMVQPLITTLAEKGITLKIQPLALKHLSEVGYDEHMGARPLRRTLQTE 780

Query: 780 VEDHLSELLLANQVKEGQVIKIGVSKGKLKFDIA 813
           +ED LSEL+L+ ++  G  +KIG+S GKL F IA
Sbjct: 781 IEDKLSELILSRELTSGHTLKIGLSHGKLTFHIA 814
```

A related GBS gene <SEQ ID 8819> and protein <SEQ ID 8820> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -13.52
GvH: Signal Score (-7.5): -2.1
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program  count: 1 value:  -2.34 threshold:  0.0
   INTEGRAL    Likelihood = -2.34   Transmembrane   32 -  48 (  32 -  49)
   PERIPHERAL  Likelihood =  0.95     112
 modified ALOM score:   0.97

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
47.4/69.6% over 804aa
                                                          Listeria monocytogenes
  EGAD|136761| ClpC ATPase Insert characterized
  GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase Insert characterized

ORF00207(298 - 2727 of 3045)
EGAD|136761|145854(2  -  806  of  825)  ClpC  ATPase  {Listeria  monocytogenes}
GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase {Listeria monocytogenes}
%Match = 33.6
%Identity = 47.4  %Similarity = 69.6
Matches = 372  Mismatches = 229  Conservative Sub.s = 174

87        117       147       177       207       237       267       297
SFF*STPIIWKYVINDWRAYQ*TSF**FDSIIIR*RDNYRT*RKFDSGDIR**RLRRASLCY*SSYAP*IITTIR*KRIP

                                                                               M

327       357       387       417       447       477       507       537
FMSHYSIKLQEVFRLAQFQAARYESHYLESWHLLLAMVLVHDSVAGLTFAEYESEVAIEEYEAATILALGRAPKEEITNY
    :: : |:|: |:| :| |     | : |:|| :|   : :|    | ||  :: |: :      :|   :: :|
MFGRFTQRAQKVLALSQEEAMRLNHSNLGTEHILLGLVREGEGIAA--KALYELGISSEKVQQEVEGLIGHG-EKAVTTI
                 20        30        40          50        60        70

567       597       627       657       687       717       744       774
QFLEQSPALKKILKLAENISIVVGAEDVGTEHVLLAMLVNKDLLATRILXLVGFRGQDDGESV-RMVDLRKALERHAGFT
|:    |  ||:::|: : :  :|   |||||:|| ::   : :| |:|  :|       : | :::  . |
QYT---PRAKKVIELSMDEARKLGHTYVGTEHILLGLIREGEGVAARVLSNLGISLNKARQQVLQLLGGGDA--------
             90       100       110       120       130       140
```

```
804       834       864       894       924       954       984      1014
KDDIKAIYELRNPKKAKSGASFSDMMKPPSTAGDLADFTRDLSQMAVDGEIEPVIGRDKEISRMVQVLSRKTKNNPVLVG
                 :||      :   |   ||   |||: :| :  ::||||| ||| |:::||||:|||||||:|
-----------------TGAGRQTNTQATPTLDSLA---RDLTVIAREDNLDPVIGRSKEIQRVIEVLSRRTKNNPVLIG
                 150       160          170       180       190       200

1044      1074      1104      1134      1164      1194      1224      1254
DAGVGKTALAYGLAQRIANGNIPYELRDMRVLELDMMSVVAGTRFRGDFEERMNQIIADIEEDGHIILFIDELHTIMGSG
: ||||||:| ||||:|    :|  ||  ||: ||| :|||||::||:||:|: ::: :| : |::|||||||||::|:|
EPGVGKTAIAEGLAQQIVRNEVPETLRGKRVMTLDMGTVVAGTKYRGEFEDRLKKVMDEIRQAGNVILFIDELHTLIGAG
          220       230       240       250       260       270       280

1284      1314      1344      1374      1404      1434      1464      1494
SGIDSTLDAANILKPALARGTLRTVGATTQEEYQKHIEKDAALSRRFAKVLVEEPNLEDAYEILLGLKPAYEAFHNVTIS
 | :  :||:||||| |||| |: :|||| :||:|:|||| || |||  : |:|| :|:: :|| ||:  ||| | | |:
-GAEGAIDASNILKPPLARGELQCIGATTLDEYRKYIEKDRALERRFQPIKVDEPTVEESIQILHGLRDRYEAHHRVAIT
           300       310       320       330       340       350       360

1524      1554      1584                1614      1644      1674      1704
DEAVMTAVKVAHRYLTSKNLPDSAIDLLDEASATVQM----------MIKKNAPSLLTEVDQAILDDDMKSASKALKASY
|||:  ||::: ||:: : ||| |||::||: : |::           :: |   |  | | |:   : : |:
DEALEAAVRLSDRYISDRFLPDKAIDVIDESGSKVRLKSFTTPKNVKEMENNLSDLKKEKDAAVQGQEFEKAASLRDKEQ
           380       390       400       410       420       430       440

1725                1737      1767      1797      1827      1857      1887
KGKK---RKPIA----------------VTEDHIMATLSRLSGIPVEKLTQADSKKYLNLEKELHKRVIGQDDAVTAISR
| ||   :| :                 |||| :   ::  :|||| || : :: | ||:|| ||:|||||| || |:|
KLKKSLDKKSLEETKANWQEKQGLDHSEVTEDIVAEVVASWTGIPVAKLAETETNKLLNMEKLLHERVIGQDAAVKAVSL
           460       470       480       490       500       510       520

1917      1947      1977      2007      2037      2067      2097      2127
AIRRNQSGIRTGKRPIGSFMFLGPTGVGKTELAKALAEVLFDDESALIRFDMSEYMEKFAASHLNGAPPGYVGYDEGGEL
|:|| ::|::  |||||||:|||||||||||||:|||| :| || ::|| |||||||||: : | |||||||||:|||:|
AVRRARAGLKDPKRPIGSFIFLGPTGVGKTELARALAESMFGDEDSMIRIDMSEYMEKFSTARLVGAPPGYVGYEEGGQL
           540       550       560       570       580       590       600

2157      2187      2217      2247      2277      2307      2337      2367
TEKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDGVLTDSRGRKVDFSNTIIIMTSNLGATALRDDKTVGFGAKDISHDY
||||| |||||:|:||:||||||:||:|||||||| ||||:|| ||| ||:||||||:||  :: ||::||   |   |:
TEKVRQKPYSVVLLDEIEKAHPDVFNMLLQVLDDGRLTDSKGRVVDFRNTVIIMTSNIGAQEMKQDKSMGFNVTDPLKDH
           620       630       640       650       660       670       680

2397      2427      2457      2487      2517      2547      2577      2607
TAMQKRIMEELKKAYRPEFINRIDEKVVFHSLSQDNMREVVKIMVKPLILALKDKGMDLKFQPSALKHLAEDGYDIEMGA
 ||: |::::||:|:|||||||||| :||||| :  ::::| ::   |   | :: : :|:   |   :|:|||| | ||
KAMEHRVLQDLKQAFRPEFINRIDETIVFHSLQEKELKQIVTLLTAQLTKRLAERDIHVKLTEGAKSKIAKDGYDPEYGA
           700       710       720       730       740       750       760

2637      2667      2697      2727      2757      2787      2817      2847
RPLRRTIQTQVEDHLSELLLANQVKEGQVIKIGVSKGKLKFDIAKS*NIPVPMGTGILI*KENVQNILDIFL*IYEK*KD
|||:| || :||| ||| ||   :| |  ::||| |||:     :
RPLKRAIQKEVEDMLSEELLRGNIKVGDYVEIGVKDGKLEVRKKDAPKKKTTSKKVKAK
           780       790       800       810       820
```

There is also homology to SEQ ID 258.

SEQ ID 8820 (GBS26) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 9; MW 93.3kDa), in Figure 167 (lane 16 & 17; MW 108kDa) and in Figure 239 (lane 14; MW 108kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 7; MW 118kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1477

A DNA sequence (GBSx1563) was identified in *S.agalactiae* <SEQ ID 4539> which encodes the amino acid sequence <SEQ ID 4540>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4541> which encodes the amino acid sequence <SEQ ID 4542>. Analysis of this protein sequence reveals the following:

```
     Possible site: 17

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 178/213 (83%), Positives = 199/213 (92%)

Query: 1   MLIVLAGTIGAGKSSLAAALGQHLGTDVFYEAVDNNPVLDLYYQDPQKYAFLLQIFFLNK 60
            MLIVLAGTIGAGKSSLAAALG+HLGTDVFYEAVDNNPVLDLYYQDP+KYAFLLQI+FLNK
Sbjct: 1   MLIVLAGTIGAGKSSLAAALGEHLGTDVFYEAVDNNPVLDLYYQDPKKYAFLLQIYFLNK 60

Query: 61  RFQSIKEAYKANNNVLDRSIFEDELFLTLNYKNGNVTKTELDIYKELLANMLEELEGMPK 120
            RF+SIKEAY+A+NN+LDRSIFEDELFL LNYKNGNVTKTELDIY+ELLANMLEELEGMPK
Sbjct: 61  RFKSIKEAYQADNNILDRSIFEDELFLKLNYKNGNVTKTELDIYQELLANMLEELEGMPK 120

Query: 121 KRPDLLVYIDVSFDKMLERIDKRGRSFEQVDSNPELYDYYKQVHSEYPEWYENYDVSPKI 180
            KRPDLL+YIDVSFDKMLERI++RGRSFEQVD NP L  YY QVH EYP WYE+Y+VSPK+
Sbjct: 121 KRPDLLIYIDVSFDKMLERIERRGRSFEQVDGNPSLEQYYHQVHGEYPTWYEDYEVSPKM 180

Query: 181 RIDGNKLDFVKNPEDLQHVLDTIDSELQKLDLL 213
            +IDGN LDFV+NP+DL  VL  ID++L++L LL
Sbjct: 181 KIDGNSLDFVQNPQDLATVLKMIDTKLKELHLL 213
```

A related GBS gene <SEQ ID 8821> and protein <SEQ ID 8822> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      3.94
GvH: Signal Score (-7.5): 1.42
     Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 0 value:   7.69 threshold:  0.0
   PERIPHERAL  Likelihood =  7.69      49
 modified ALOM score:  -2.04

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 4540 (GBS9) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 5; MW 52kDa) and Figure 12 (lane 2 & 3; MW 50.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 6; MW 27kDa) and Figure 3 (lane 2; MW 25kDa). The GBS9-GST fusion product was purified (Figure 191, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 318), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1478

A DNA sequence (GBSx1564) was identified in *S.agalactiae* <SEQ ID 4543> which encodes the amino acid sequence <SEQ ID 4544>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1182(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4545> which encodes the amino acid sequence <SEQ ID 4546>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 281/323 (86%), Positives = 305/323 (93%)

Query: 3   QLNSSFMIGKVEIPHRTVLAPMAGITNSAFRTIAKEFGAGLVVMEMISEKGLLYNNEKTL 62
            +LNSSF IG VEIPHRTVLAPMAG+TNSAFRTIAKEFGAGLVVMEMISEKGLLYNNEKTL
Sbjct: 27  KLNSSFRIGDVEIPHRTVLAPMAGVTNSAFRTIAKEFGAGLVVMEMISEKGLLYNNEKTL 86

Query: 63  HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQSNTKADIVDINMGCPVNKVVKNEAGAKW 122
            HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQ+NTKADIVDINMGCPVNKVVKNEAGAKW
Sbjct: 87  HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQTNTKADIVDINMGCPVNKVVKNEAGAKW 146

Query: 123 LRDPEKIYHIVKEVTSVLDIPLTVKMRTGWSDSSNAIENALAAESAGVSALAMHGRTREQ 182
            LRDP+KIYHIVKEVTSVLDIPLTVKMRTGW+DSS A+ENALAAESAGVSALAMHGRTREQ
Sbjct: 147 LRDPDKIYHIVKEVTSVLDIPLTVKMRTGWADSSLAVENALAAESAGVSALAMHGRTREQ 206

Query: 183 MYTGTCDHETLGKVAKAVTSIPFIANGDIRTVHDAKFMIEEIGADAIMVGRGARSNPYIF 242
            MYTGTCDHETL +V+KA+T IPFI NGD+R+V DAKFMIEEIG DA+M+GR A +NPY+F
Sbjct: 207 MYTGTCDHETLARVSKAITKIPFIGNGDVRSVQDAKFMIEEIGVDAVMIGRAAMNNPYLF 266

Query: 243 TQINHFFETGEILPDLPFEKMLDVAEDHLTRLVNLKGETIAVREFRGLAPHYLRGKSGAA 302
            TQINHFFETG+ LPDLPF K LD+A+DHL RL+NLKGETIAVREFRGLAPHYLRG +GAA
Sbjct: 267 TQINHFFETGQELPDLPFAKKLDIAKDHLKRLINLKGETIAVREFRGLAPHYLRGTAGAA 326

Query: 303 KIRGAVSRAETLAEVQELFAGLR 325
            K+RGAVSRAETLAEV+ +F  +R
Sbjct: 327 KVRGAVSRAETLAEVEAIFETVR 349
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1479

A DNA sequence (GBSx1565) was identified in *S.agalactiae* <SEQ ID 4547> which encodes the amino acid sequence <SEQ ID 4548>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2164(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 3930:

```
Identities = 235/288 (81%), Positives = 259/288 (89%)

Query: 1   MDKIIKSISTSGSFRAYVLDCTETVRTAQEKHQTLSSSTVALGRTLIANQILAANQKGNS 60
            MDKIIKSI+ SG+FRAYVLD TETV  AQEKH TLSSSTVALGRTLIANQILAANQKG+S
Sbjct: 1   MDKIIKSIAQSGAFRAYVLDSTETVALAQEKHNTLSSSTVALGRTLIANQILAANQKGDS 60

Query: 61  KVTVKVIGDSSFGHIISVADTKGNVKGYIQNTGVDIKKTATGEVLVGPFMGNGHFVVITD 120
            K+TVKVIGDSSFGHIISVADTKG+VKGYIQNTGVDIKKTATGEVLVGPFMGNGHFV I D
Sbjct: 61  KITVKVIGDSSFGHIISVADTKGHVKGYIQNTGVDIKKTATGEVLVGPFMGNGHFVTIID 120

Query: 121 YATGQPYTSTTPLITGEIGEDFAYYLTESEQTPSAVGLNVLLDDEDKVKVAGGFMLQVLP 180
            Y TG PYTSTTPLITGEIGEDFAYYLTESEQTPSA+GLNVLLD+ DKVKVAGGFM+QVLP
Sbjct: 121 YGTGNPYTSTTPLITGEIGEDFAYYLTESEQTPSAIGLNVLLDENDKVKVAGGFMVQVLP 180

Query: 181 GASDEEISRYEKRIQEMPSISSLLESENHIESLLSAIYGEDDYKRLSEDSLAFYCDCSKE 240
            GAS+EEI+RYEKR+QEMP+IS LL S+NH+++LL AIYG++ YKRLSE+ L+F CDCS+E
Sbjct: 181 GASEEEIARYEKRLQEMPAISHLLASKNHVDALLEAIYGDEPYKRLSEEPLSFQCDCSRE 240

Query: 241 RFEAALLTLGTKELQAMKDEDKGVEITCQFCNQTYYFTEEDLEKIIND 288
            RFEAAL+TL   +LQAM DEDKG EI CQFC   Y F E DLE II+D
Sbjct: 241 RFEAALMTLPKADLQAMIDEDKGAEIVCQFCGTKYQFNESDLEAIISD 288
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1480

A DNA sequence (GBSx1566) was identified in *S.agalactiae* <SEQ ID 4549> which encodes the amino acid sequence <SEQ ID 4550>. This protein is predicted to be surface-located membrane protein 1 (lmp1). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4312(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB93480 GB:AF019377 tellurite resistance protein [Rhodobacter
           sphaeroides]
 Identities = 64/350 (18%), Positives = 146/350 (41%), Gaps = 7/350 (2%)
```

```
Query: 44  LTPAQKSAISEKTPALVDTFVGDQNALLDFGQSAVEGVNTTVNHILSEQKKIQIPQVDDL 103
           L  A      E    + +  V D  +++ FG  A   + T    +L++ K   +    D
Sbjct: 34  LASAPPEKAQEIRRRMAELNVSDSQSIIGFGSKAQAELQTISQQMLADVKNKDVGPAGDS 93

Query: 104 LKNANRELNGFIAKYKDATPAELEKKPNLIQKLFKQSKTSLQEFYFDSQNIEQKMDMMAA 163
           L+     + GF       +  ++ +K +  ++L  ++      F    ++++Q++D +
Sbjct: 94  LREVVSTIRGF-----SVSEFDVRRKASWWERLLGRT-APFARFVARYEDVQQQIDRITQ 147

Query: 164 NVVKQEDTLARNIVSAEMLIEDNTKSIENLVGVIAFIESSQAEAANRASHLQQEILALDS 223
           +++  E  L ++I   ++L        + L   IA  +   A+   R    ++  +A
Sbjct: 148 SLLTHEHRLLKDIKGLDILYARTLDFYDELALYIAAGDEVLADLDGRVIPAKEAEVAATP 207

Query: 224 QTSEYQIKSNQLARMTEVINTLEQQHPEYVSRLYVAWATTPQMRNLVKVSSDMRQKLGML 283
           +  +  IK+ +L  +    + LE++  +      V   + P +R + +    +  ++
Sbjct: 208 E-GDRMIKAQELRDLRAARDDLERRVHDLKLTRQVTMQSLPSIRLVQENDKALVTRINST 266

Query: 284 RRNTIPTMKLSIAQLGMMQQSVKSGVTADAIVNANNAALQMLAETSKEAIPMLEKTAQSP 343
             NT+P  +  +AQ   +Q+S ++        +  N  L   AE  ++A  ++ K  +
Sbjct: 267 LVNTVPLWETQLAQAVTIQRSREAAEAVRGASDLTNELLTANAENLQQANKIVRKEMERG 326

Query: 344 TVSIKSVTALAESLVAQNNGIIAAIDKGRKERAQLESAVIKSAETINDSV 393
              I++V     +L+A  N  +A  D+GR  RA  E+ + +    + D++
Sbjct: 327 VFDIEAVKKANATLIATINESLAIADEGRARRATAETELQRMEAELRDTL 376
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4551> which encodes the amino acid sequence <SEQ ID 4552>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3230(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 333/413 (80%), Positives = 379/413 (91%)

Query: 5   FNFDIDQIADNAITKTDKTTEIISNQTTSQTGQIAFFEKLTPAQKSAISEKTPALVDTFV 64
           FNFDIDQIADNA+ KTDKTT+IIS+  T   GQI+FFEKL+  Q++AI+ K PALVDTF+
Sbjct: 4   FNFDIDQIADNAVIKTDKTTDIISDLPTDTNGQISFFEKLSADQQTAITAKAPALVDTFL 63

Query: 65  GDQNALLDFGQSAVEGVNTTVNHILSEQKKIQIPQVDDLLKNANRELNGFIAKYKDATPA 124
            DQNALLDFGQSAVEGVN TVNHIL+EQKK+QIPQVDDLLK+ NRELNGFIAKYKDATP
Sbjct: 64  ADQNALLDFGQSAVEGVNATVNHILAEQKKLQIPQVDDLLKSTNRELNGFIAKYKDATPV 123

Query: 125 ELEKKPNLIQKLFKQSKTSLQEFYFDSQNIEQKMDMMAANVVKQEDTLARNIVSAEMLIE 184
           +L+KKPN +QKLFKQS+ +LQEFYFDSQNIEQKMD MAA VVKQEDTLARNIVSAE+LIE
Sbjct: 124 DLDKKPNFLQKLFKQSRDTLQEFYFDSQNIEQKMDSMAAAVVKQEDTLARNIVSAELLIE 183

Query: 185 DNTKSIENLVGVIAFIESSQAEAANRASHLQQEILALDSQTSEYQIKSNQLARMTEVINT 244
           DNTKSIE+LVGVIAFIE+SQ EA+ RA+ LQ+++   DS T +YQIK++ LAR TEVINT
Sbjct: 184 DNTKSIEHLVGVIAFIEASQKEASQRAAALQKDLKTKDSATPDYQIKADLLARTTEVINT 243

Query: 245 LEQQHPEYVSRLYVAWATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS 304
           LEQQH EY+SRLYVAWATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS
Sbjct: 244 LEQQHTEYLSRLYVAWATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS 303

Query: 305 VKSGVTADAIVNANNAALQMLAETSKEAIPMLEKTAQSPTVSIKSVTALAESLVAQNNGI 364
           VKSG+TADAI+NANNAALQMLAETSKEAIP LE++AQ+PT+S+KSVT+LAESLVAQNNGI
Sbjct: 304 VKSGMTADAIINANNAALQMLAETSKEAIPALEQSAQNPTLSMKSVTSLAESLVAQNNGI 363

Query: 365 IAAIDKGRKERAQLESAVIKSAETINDSVKIRDKKIVEALLNEGKSTQEKVDE 417
           IAAID GRKERAQLESA+I+SAETINDSVK+RD+ IV+ALL+EGK TQ+ +D+
Sbjct: 364 IAAIDHGRKERAQLESAIIRSAETINDSVKLRDQNIVQALLSEGKETQKTIDK 416
```

SEQ ID 4550 (GBS201) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 5; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 3; MW 74.5kDa) and in Figure 62 (lane 8 & 9; MW 74.5kDa). The GBS201-GST fusion product was purified (Figure 209, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 304), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1481

A DNA sequence (GBSx1567) was identified in *S.agalactiae* <SEQ ID 4553> which encodes the amino acid sequence <SEQ ID 4554>. This protein is predicted to be rhoptry protein. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.58    Transmembrane    13 -  29 (  10 -  31)
   INTEGRAL    Likelihood = -1.54    Transmembrane    33 -  49 (  33 -  49)

----- Final Results -----
               bacterial membrane --- Certainty=0.3633(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4555> which encodes the amino acid sequence <SEQ ID 4556>. Analysis of this protein sequence reveals the following:

```
     Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 115/239 (48%), Positives = 162/239 (67%), Gaps = 3/239 (1%)

Query: 32  EVIATLLIIGGGYCAYYVYD-KKRLKRFTSNQRIEALKSDIKETDQDIRHLEILKKDNRS 90
            +++  + I G GY  + V   +KRL +    +++E LK+ I+  D+ +R L+    D+
Sbjct: 42  DILPAIAIGGTGYAIFRVRSHQKRLAKAKIAKQLEDLKAKIQLADRKVRLLDTYLADHDD 101

Query: 91  KEYIKLAHQILPQLDLIRNEANQLQKAIEPNIYKRITKKANTFSNEINEQLIKLHASPEL 150
             +Y  LA Q+LPQL  I+ +A  L+  ++P IY+RITKKAN   ++I  QL  L  + L
Sbjct: 102 FQYNVLAQQLLPQLSDIKAKAITLKDQLDPQIYRRITKKANDVESDITLQLETLQIATTL 161

Query: 151 --EPISDQEDEMIRIAPELKPFYHNIQDDHFAILKKIEEADNKAELAAIHQANMKRFTDV 208
              +P+      +I  APELKP+Y NIQ DH AIL KI+ ADN+ EL A+H ANM+RF D+
Sbjct: 162 NPQPLKTPSPNLINKAPELKPYYDNIQTDHQAILAKIQGADNQEELLALHDANMRRFEDI 221

Query: 209 LAGYIRIKQSPKNFNNAKERLEQALQAIKKFNLDLDETLRQLNESDMKDFDVSLRMMQG 267
            L GY++IK+ PKN+ NA  RLEQA QAI++F+ DLDETLR+LNESD+KDFD+SLR+MQG
Sbjct: 222 LTGYLKIKEEPKNYYNAAARLEQAKQAIQQFDEDLDETLRRLNESDLKDFDISLRIMQG 280
```

SEQ ID 4554 (GBS265) was expressed in *E.coli* as GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 2; MW 56kDa) and in Figure 62 (lane 6; MW 56.3kDa).

The GBS265-GST fusion product was purified (Figure 207, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 258A) and FACS (Figure 258B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1482

A DNA sequence (GBSx1568) was identified in *S.agalactiae* <SEQ ID 4557> which encodes the amino acid sequence <SEQ ID 4558>. This protein is predicted to be glutamate--cysteine ligase (gshA). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane  575 - 591 ( 575 - 591)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG08588 GB:AE004933 glutamate--cysteine ligase [Pseudomonas aeruginosa]
 Identities = 142/468 (30%), Positives = 220/468 (46%), Gaps = 62/468 (13%)

Query: 12  SHLPIL-QATFGLERESLRIHQPTQRVAQTPHPKTLGSRNYHPYIQTDYSEPQLELITPI 70
            ++LP+L +   G+ERE LR+     ++A TPHP+ LGS   HP I TDYSE  LE ITP
Sbjct: 16  ANLPLLTECLHGIERECLRVDSDG-KLALTPHPRALGSTLTHPQITTDYSEALLEFITPT 74

Query: 71  AKDSQEAIRFLKAISDVAGRSINHDEYLWPLSMPPKV-REEDIQIAQLEDA----FEYDY 125
              D  + +  L+ I   A   ++  EYLW  SMP ++  EE I IA+   +     +Y Y
Sbjct: 75  ETDVADTLGDLERIHRFASSKLD-GEYLWSPSMPCELPDEESIPIARYGSSMIGRLKYVY 133

Query: 126 RKYLEKTYGKLIQSISGIHYNLGLGQELLTSLFELSQAD-NAIDFQNQLYMKLSQNFLRY 184
            RK L   YGK +Q I+GIHYN  L + L   L +   ++ +  D+Q+  Y+ L +NF RY
Sbjct: 134 RKGLALRYGKTMQCIAGIHYNFSLPERLWPLLRQAEGSELSERDYQSAAYIALIRNFRRY 193

Query: 185 RWLLTYLYGASPVAEEDFLDQKLNNPVR------------SLRNSHLGYVNHKDIRIS-- 230
             WLL YL+GASP  +  FL  + +   R            SLR S LGY N+    ++
Sbjct: 194 SWLLMYLFGASPALDAGFLRGRPSQLERLDEHTLYLPYATSLRMSDLGYQNNAQAGLTPC 253

Query: 231 YTSLKDYVNDLENAV---------------------KSGQLIAEKEFYSPVRLR-----G 264
            Y  L+ Y++ L  AV                      +  L  E E+YS +R +     G
Sbjct: 254 YNDLQSYIDSLRQAVSTPYPPYEKVGTKQDGEWVQLNTNILQIENEYYSSIRPKRVTYTG 313

Query: 265 SKACRNYLEKGITYLEFRTFDLNPFSPIGITQETVDTVHLFLLALLWIDS---------- 314
             +  +    +G+ Y+E R  D+NPF P+GI  +    +  FLL   + DS
Sbjct: 314 ERPVQALAARGVQYVEVRCLDINPFLPLGIDLDEARFLDAFLLFCAFSDSPLLNGECSDA 373

Query: 315 SSHIDQDIKEANRLN-DLIALSHPLEKLPNQAPVSDLVDAMQSVIQHFNLSPYYQDLLES 373
           + +    +KE  R    L     P+E       + + +    +++        +   L +
Sbjct: 374 TDNFLAVVKEGRRPGLQLQRRGQPVELQVWANELLERIADTAALLDRARGGEAHAAALAA 433

Query: 374 VKRQIQSPELTVAGQLLEMI--EGLSLETFGQRQGQIYHDYAWEAPYA 419
            + ++   ELT + Q+L+++   G S E F  RQ + + +Y  + P A
Sbjct: 434 QRAKVADAELTPSAQVLKVMRERGESFEAFSLRQSREHAEYFRQHPLA 481
```

There is also homology to SEQ ID 4560.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1483

A DNA sequence (GBSx1569) was identified in *S.agalactiae* <SEQ ID 4561> which encodes the amino acid sequence <SEQ ID 4562>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1504(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB73814 GB:AL139078 helix-turn-helix containing protein
           [Campylobacter jejuni]
 Identities = 107/223 (47%), Positives = 148/223 (65%), Gaps = 7/223 (3%)

Query: 1   MDKEKLDYWKTIITFLHNVLGDNYEIVLHVVDENDIYIGELVNSHISGRTISSPLTTFAL 60
           MD+ +   +  +  FL  VLG+ YEIV HV+ E+  YI  + NSHISGR++ SPLT FA
Sbjct: 1   MDEGQKQQFIKLTYFLGEVLGEQYEIVFHVITEDGAYIAAIANSHISGRSLDSPLTAFAS 60

Query: 61  DLIKNKVYKEKDFVTNYKAIVSPLNKEVRGSTFFIKNAQNELEGMLCINLDISAYQNIAL 120
           +L++NK Y EKDF+ +YKA+V   +K +RGSTFFIKN  ++L G+LCIN D S  +++
Sbjct: 61  ELMQNKKYLEKDFLCDYKALVGK-SKLIRGSTFFIKN-HDKLVGILCINHDTSIMRDLIC 118

Query: 121 DILDLVNL-NVNKILPKSPQKISLPQQEEPVEVLSGNIQDIISEIVDPSLLNQNIHLSQE 179
            ++DL  + ++  IL      IS  Q +  +E LS +I+DI+ + VD S LN +  LS
Sbjct: 119 KMIDLEKIGDMGDIL----GNISFSQNDSSIETLSHSIEDILVQSVDSSYLNSDYQLSIT 174

Query: 180 VKVEIVSKLHEKGVFQLKGAVSKVAEVLNISEPSVYRYLKKIE 222
            K EI  KL+EKG+F +KGAV  VA+ L ISEPSVYRYLKK +
Sbjct: 175 QKEEIAEKLYEKGIFNIKGAVPIVAKFLKISEPSVYRYLKKFK 217
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4563> which encodes the amino acid sequence <SEQ ID 4564>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1636(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 169/224 (75%), Positives = 198/224 (87%), Gaps = 3/224 (1%)

Query: 1   MDKEKLDYWKTIITFLHNVLGDNYEIVLHVVDENDIYIGELVNSHISGRTISSPLTTFAL 60
           MDKE L+YWKT+ITFLH+VLGDNYEI+LHV+D+NDIYIGELVNSHISGR+  SPLTTFAL
Sbjct: 1   MDKETLNYWKTVITFLHDVLGDNYEIILHVIDKNDIYIGELVNSHISGRSKQSPLTTFAL 60

Query: 61  DLIKNKVYKEKDFVTNYKAIVSPLNKEVRGSTFFIKNAQNELEGMLCINLDISAYQNIAL 120
           DLI NKVYKEKDFVTNYKAIVSP +KEVRGSTFFIK+ +  LEGMLCINLDISAYQ +A
Sbjct: 61  DLITNKVYKEKDFVTNYKAIVSPQHKEVRGSTFFIKDKKGNLEGMLCINLDISAYQGVAR 120

Query: 121 DILDLVNLNVNKILP--KSPQKISLPQQEEPVEVLSGNIQDIISEIVDPSLLNQNIHLSQ 178
           D+L LVNLN+   +P  K P+ ++ PQ EE VE+L+ NIQDII +I+DPSLL  N+HLSQ
```

```
Sbjct: 121 DLLKLVNLNLEHFIPTAKEPKTVT-PQPEEAVEILTSNIQDIIGQIIDPSLLRHNVHLSQ 179

Query: 179 EVKVEIVSKLHEKGVFQLKGAVSKVAEVLNISEPSVYRYLKKIE 222
           +VK++IV+KL+EKGVFQLKGAVSKVA++L ISEPSVYRYLKKIE
Sbjct: 180 DVKIDIVAKLYEKGVFQLKGAVSKVADILCISEPSVYRYLKKIE 223
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1484

A DNA sequence (GBSx1570) was identified in *S.agalactiae* <SEQ ID 4565> which encodes the amino acid sequence <SEQ ID 4566>. This protein is predicted to be regulatory protein pfoR. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.80    Transmembrane  299 - 315 ( 296 - 325)
   INTEGRAL    Likelihood = -7.54    Transmembrane  172 - 188 ( 169 - 193)
   INTEGRAL    Likelihood = -7.17    Transmembrane   71 -  87 (  66 -  98)
   INTEGRAL    Likelihood = -4.99    Transmembrane  261 - 277 ( 260 - 278)
   INTEGRAL    Likelihood = -2.81    Transmembrane  128 - 144 ( 127 - 149)
   INTEGRAL    Likelihood = -2.18    Transmembrane  101 - 117 ( 101 - 119)
   INTEGRAL    Likelihood = -0.53    Transmembrane  198 - 214 ( 197 - 214)

----- Final Results -----
            bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA60239 GB:X86525 pfoS [Clostridium perfringens]
 Identities = 96/147 (65%), Positives = 122/147 (82%)

Query: 100 GTGIIPGFLAGYLVGFLVKWMERNIPGGLDLISIIIIGAPLTRLVAKLLTPLINSTLLTI 159
           G GI+PGF+AGYL  F++K++E+ IP GLDLI II++GAPL R +A +  PL+ +TL  I
Sbjct: 1   GFGILPGFIAGYLGSFVIKFLEKKIPAGLDLIVIIVLGAPLVRGIAAISNPLVETTLQNI 60

Query: 160 GDILTSGAHSNPILMGIILGGTIVVVATAPLSSMALTAMLGLTGMPMAIGALSVFGSSFM 219
           G ++T+ + ++PI+MGIILGG + VVATAPLSSMALTAMLGLTG+PMAIGAL+VFGSSFM
Sbjct: 61  GGVITATSTASPIMMGIILGGIVTVVATAPLSSMALTAMLGLTGLPMAIGALAVFGSSFM 120

Query: 220 NGVLFHKLKLGSRKDNIAFAVEPLTQA 246
           N V F K+K GS+KD IA A+EPLTQA
Sbjct: 121 NLVFFGKMKFGSKKDTIAVAIEPLTQA 147
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4567> which encodes the amino acid sequence <SEQ ID 4568>. Analysis of this protein sequence reveals the following:

```
     Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.70    Transmembrane  303 - 319 ( 296 - 325)
   INTEGRAL    Likelihood = -7.11    Transmembrane   70 -  86 (  66 -  98)
   INTEGRAL    Likelihood = -6.53    Transmembrane  172 - 188 ( 169 - 193)
   INTEGRAL    Likelihood = -4.83    Transmembrane  261 - 277 ( 260 - 278)
   INTEGRAL    Likelihood = -2.55    Transmembrane  101 - 117 ( 101 - 119)
   INTEGRAL    Likelihood = -2.28    Transmembrane  124 - 140 ( 124 - 140)
   INTEGRAL    Likelihood = -1.91    Transmembrane  198 - 214 ( 197 - 215)

----- Final Results -----
            bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA60239 GB:X86525 pfoS [Clostridium perfringens]
 Identities = 95/147 (64%), Positives = 123/147 (83%)

Query: 100 GTGIIPGFVAGYVVSFLIKWMEKNIPGGLDLISIIIVGAPLTRFLAQLITPVINSTLLTI 159
           G GI+PGF+AGY+ SF+IK++EK IP GLDLI II++GAPL R +A +  P++ +TL  I
Sbjct: 1   GFGILPGFIAGYLGSFVIKFLEKKIPAGLDLIVIIVLGAPLVRGIAAISNPLVETTLQNI 60

Query: 160 GDILTSSANSNPIIMGMILGGTIVVVATAPLSSMALTAMLGLTGIPMAIGALSVFGSSFM 219
           G ++T+++ ++PI+MG+ILGG + VVATAPLSSMALTAMLGLTG+PMAIGAL+VFGSSFM
Sbjct: 61  GGVITATSTASPIMMGIILGGIVTVVATAPLSSMALTAMLGLTGLPMAIGALAVFGSSFM 120

Query: 220 NGVLFYRLKLGERKDNIAFAIEPLTQA 246
           N V F ++K G +KD IA AIEPLTQA
Sbjct: 121 NLVFFGKMKFGSKKDTIAVAIEPLTQA 147
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 302/339 (89%), Positives = 330/339 (97%)

Query: 1   MNIIIGTSLLILVLAIFTLFNYKAPYGTKAMGALASAACASFLVEAFQDSFFGKVLGFQF 60
           M+IIIGTSLLILVLAIF+LFNYKAP+G KAMGALASAACASFLVEAFQDSFFGKVLGFQF
Sbjct: 1   MDIIIGTSLLILVLAIFSLFNYKAPHGAKAMGALASAACASFLVEAFQDSFFGKVLGFQF 60

Query: 61  LSEVGGANGSLSGVAAAILVAIAIGVTPGYAVLIGLSVSGTGIIPGFLAGYLVGFLVKWM 120
           LSEVGGANGSLSGVAAAILVAIAIGV+PGYAVLIGLSVSGTGIIPGF+AGY+V FL+KWM
Sbjct: 61  LSEVGGANGSLSGVAAAILVAIAIGVSPGYAVLIGLSVSGTGIIPGFVAGYVVSFLIKWM 120

Query: 121 ERNIPGGLDLISIIIIGAPLTRLVAKLLTPLINSTLLTIGDILTSGAHSNPILMGIILGG 180
           E+NIPGGLDLISIII+GAPLTR +A+L+TP+INSTLLTIGDILTS A+SNPI+MG+ILGG
Sbjct: 121 EKNIPGGLDLISIIIVGAPLTRFLAQLITPVINSTLLTIGDILTSSANSNPIIMGMILGG 180

Query: 181 TIVVVATAPLSSMALTAMLGLTGMPMAIGALSVFGSSFMNGVLFHKLKLGSRKDNIAFAV 240
           TIVVVATAPLSSMALTAMLGLTG+PMAIGALSVFGSSFMNGVLF++LKLG RKDNIAFA+
Sbjct: 181 TIVVVATAPLSSMALTAMLGLTGIPMAIGALSVFGSSFMNGVLFYRLKLGERKDNIAFAI 240

Query: 241 EPLTQADVTSANPIPIYVTNFVGGAACGILIALMKLVNDTPGTATPIAGFAVMFAYNPMI 300
           EPLTQADVTSANPIPIYVTNFVGGAACG+LIALMKLVNDTPGTATPIAGFAVMFAYNP+
Sbjct: 241 EPLTQADVTSANPIPIYVTNFVGGAACGVLIALMKLVNDTPGTATPIAGFAVMFAYNPVA 300

Query: 301 KVLITALGCIILSLLAGYFGGIVFKDYKLVTKEELQARD 339
           KVLITALGCII+SL+ GY GG VFK+Y+LVTK+ELQAR+
Sbjct: 301 KVLITALGCIIISLIVGYIGGSVFKNYRLVTKQELQARN 339
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1485

A DNA sequence (GBSx1571) was identified in *S.agalactiae* <SEQ ID 4569> which encodes the amino acid sequence <SEQ ID 4570>. This protein is predicted to be adenylosuccinate synthetase (purA). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0560(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB16079 GB:Z99124 adenylosuccinate synthetase [Bacillus subtilis]
 Identities = 320/427 (74%), Positives = 378/427 (87%)

Query: 1   MTSVVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVIDNKKFKLHLIPSGIF 60
           M+SVVVVGTQWGDEGKGKITDFLS +AEVIARYQGG+NAGHTI  D   +KLHLIPSGIF
Sbjct: 1   MSSVVVVGTQWGDEGKGKITDFLSENAEVIARYQGGNNAGHTIKFDGITYKLHLIPSGIF 60

Query: 61  FKEKISVIGNGVVVNPKSLVKELAYLHGEGVTTDNLRISDRAHVILPYHIKLDQLQEDAK 120
           +K+K  VIGNG+VV+PK+LV ELAYLH   V+TDNLRIS+RAHVILPYH+KLD+++E+ K
Sbjct: 61  YKDKTCVIGNGMVVDPKALVTELAYLHERNVSTDNLRISNRAHVILPYHLKLDEVEEERK 120

Query: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDREVFAERLKINLAEKNRLFEKMYDSTPL 180
           G NKIGTT KGIGPAYMDKAAR+GIRIADLLDR+ FAE+L+ NL EKNRL EKMY++
Sbjct: 121 GANKIGTTKKGIGPAYMDKAARIGIRIADLLDRDAFAEKLERNLEEKNRLLEKMYETEGF 180

Query: 181 EFDDIFEEYYEYGQQIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240
           + +DI +EYYEYGQQIK+YV DTSV+LNDALD G+RVLFEGAQGVMLDIDQGTYPFVTSS
Sbjct: 181 KLEDILDEYYEYGQQIKKYVCDTSVVLNDALDEGRRVLFEGAQGVMLDIDQGTYPFVTSS 240

Query: 241 NPVAGGVTIGSGVGPSKINKVVGVCKAYTSRVGDGPFPTELFDEVGDRIREIGKEYGTTT 300
           NPVAGGVTIGSGVGP+KI  VVGV KAYT+RVGDGPFPTEL DE+GD+IRE+G+EYGTTT
Sbjct: 241 NPVAGGVTIGSGVGPTKIKHVVGVSKAYTTRVGDGPFPTELKDEIGDQIREVGREYGTTT 300

Query: 301 GRPRRVGWFDSVVMRHSRRVSGITNLSLNSIDVLSGLDTVKICVAYDLDGKRIDYYPASL 360
           GRPRRVGWFDSVV+RH+RRVSGIT+LSLNSIDVL+G++T+KICVAY   G+ I+ +PASL
Sbjct: 301 GRPRRVGWFDSVVVRHARRVSGITDLSLNSIDVLAGIETLKICVAYRYKGEIIEEFPASL 360

Query: 361 EQLKRCKPIYEELPGWSEDITACRSLDDLPENARNYVRRVGELVGVRISTFSVGPGREQT 420
           + L  C+P+YEE+PGW+EDIT  +SL +LPENAR+Y+ RV +L G+ +S FSVGP R QT
Sbjct: 361 KALAECEPVYEEMPGWTEDITGAKSLSELPENARHYLERVSQLTGIPLSIFSVGPDRSQT 420

Query: 421 NILESVW 427
           N+L SV+
Sbjct: 421 NVLRSVY 427
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4571> which encodes the amino acid sequence <SEQ ID 4572>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0560(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 406/430 (94%), Positives = 421/430 (97%)

Query: 1   MTSVVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVIDNKKFKLHLIPSGIF 60
           MTSVVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVID KKFKLHLIPSGIF
Sbjct: 1   MTSVVVVGTQWGDEGKGKITDFLSADAEVIARYQGGDNAGHTIVIDGKKFKLHLIPSGIF 60

Query: 61  FKEKISVIGNGVVVNPKSLVKELAYLHGEGVTTDNLRISDRAHVILPYHIKLDQLQEDAK 120
           F +KISVIGNGVVVNPKSLVKELAYLH EGVTTDNLRISDRAHVILPYHI+LDQLQEDAK
Sbjct: 61  FPQKISVIGNGVVVNPKSLVKELAYLHDEGVTTDNLRISDRAHVILPYHIQLDQLQEDAK 120

Query: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDREVFAERLKINLAEKNRLFEKMYDSTPL 180
           GDNKIGTTIKGIGPAYMDKAARVGIRIADLLD+++FAERL+INLAEKNRLFEKMYDSTPL
Sbjct: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDKDIFAERLRINLAEKNRLFEKMYDSTPL 180

Query: 181 EFDDIFEEYYEYGQQIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240
           +FD IFEEYY YGQ+IKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS
Sbjct: 181 DFDAIFEEYYAYGQEIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240

Query: 241 NPVAGGVTIGSGVGPSKINKVVGVCKAYTSRVGDGPFPTELFDEVGDRIREIGKEYGTTT 300
```

```
                NPVAGGVTIGSGVGP+KINKVVGVCKAYTSRVGDGPFPTELFDEVG+RIRE+G EYGTTT
Sbjct: 241 NPVAGGVTIGSGVGPNKINKVVGVCKAYTSRVGDGPFPTELFDEVGERIREVGHEYGTTT 300

Query: 301 GRPRRVGWFDSVVMRHSRRVSGITNLSLNSIDVLSGLDTVKICVAYDLDGKRIDYYPASL 360
            GRPRRVGWFDSVVMRHSRRVSGITNLSLNSIDVLSGLDTVKICVAYDLDGKRIDYYPA+L
Sbjct: 301 GRPRRVGWFDSVVMRHSRRVSGITNLSLNSIDVLSGLDTVKICVAYDLDGKRIDYYPANL 360

Query: 361 EQLKRCKPIYEELPGWSEDITACRSLDDLPENARNYVRRVGELVGVRISTFSVGPGREQT 420
            EQLKRCKPIYEELPGW EDIT  RSLD+LPENARNYVRRVGELVGVRISTFSVGPGREQT
Sbjct: 361 EQLKRCKPIYEELPGWQEDITGVRSLDELPENARNYVRRVGELVGVRISTFSVGPGREQT 420

Query: 421 NILESVWSNI 430
            NILESVW++I
Sbjct: 421 NILESVWASI 430
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1486

A DNA sequence (GBSx1572) was identified in *S.agalactiae* <SEQ ID 4573> which encodes the amino acid sequence <SEQ ID 4574>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.29   Transmembrane   30 -  46 (  22 -  55)
   INTEGRAL    Likelihood = -2.97   Transmembrane  110 - 126 ( 109 - 126)
   INTEGRAL    Likelihood = -0.11   Transmembrane   89 - 105 (  89 - 106)

----- Final Results -----
               bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8823> which encodes amino acid sequence <SEQ ID 8824> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
SRCFLG: 0
McG: Length  of  UR:   5
     Peak Value of UR:   3.05
     Net Charge of CR: 0
McG: Discrim Score:      4.64
GvH: Signal Score (-7.5): -1.66
     Possible site: 36
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 37
ALOM program   count: 2 value:  -2.97 threshold:  0.0
   INTEGRAL    Likelihood = -2.97   Transmembrane  100 - 116 (  99 - 116)
   PERIPHERAL  Likelihood =  1.38      56
 modified ALOM score:   1.09
icml HYPID: 7  CFP: 0.219

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2190(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database and no corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1487

A DNA sequence (GBSx1573) was identified in *S.agalactiae* <SEQ ID 4575> which encodes the amino acid sequence <SEQ ID 4576>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0967(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1488

A DNA sequence (GBSx1574) was identified in *S.agalactiae* <SEQ ID 4577> which encodes the amino acid sequence <SEQ ID 4578>. This protein is predicted to be SgaT protein (sgaT). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.80   Transmembrane  441 - 457 ( 436 - 464)
   INTEGRAL    Likelihood = -7.64   Transmembrane  344 - 360 ( 339 - 376)
   INTEGRAL    Likelihood = -6.58   Transmembrane  403 - 419 ( 392 - 422)
   INTEGRAL    Likelihood = -6.48   Transmembrane  237 - 253 ( 235 - 261)
   INTEGRAL    Likelihood = -5.79   Transmembrane  105 - 121 (  99 - 127)
   INTEGRAL    Likelihood = -5.52   Transmembrane  138 - 154 ( 137 - 155)
   INTEGRAL    Likelihood = -4.78   Transmembrane   18 -  34 (  14 -  38)
   INTEGRAL    Likelihood = -2.97   Transmembrane  365 - 381 ( 365 - 383)
   INTEGRAL    Likelihood = -0.69   Transmembrane   41 -  57 (  41 -  57)
   INTEGRAL    Likelihood = -0.16   Transmembrane  160 - 176 ( 160 - 176)

----- Final Results -----
               bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC77150 GB:AE000491 orf, hypothetical protein [Escherichia coli K12]
 Identities = 181/451 (40%), Positives = 274/451 (60%), Gaps = 25/451 (5%)

Query: 11  FSQNILQNPAFFVGLLVLIGYLLLKKPLHDVFAGFIKATVGYLILNVGAGGLVNTFRPIL 70
            F   ++ N    +G++  +GY+LL+K +  +  G IK  +G+++L  G+G L +TF+P++
Sbjct: 30  FFNQVMTNAPLLLGIVTCLGYILLRKSVSVIIKGTIKTIIGFMLLQAGSGILTSTFKPVV 89

Query: 71  VALAKKFNLEAAVIDPYFGLASANAKLETMG-FISVATTALLIGFGINILLVALRKVTKV 129
              +++ + +  A+ D Y   AS  A ++ MG   S    A+L+   +NI  V LR++T +
Sbjct: 90  AKMSEVYGINGAISDTY---ASMMATIDRMGDAYSWVGYAVLLALALNICYVLLRRITGI 146

Query: 130 RTLFITGHIMVQQAATISVFVLLLIPQLRNGFGAWAV----GIICGLYWAVSSNMTVEAT 185
            RT+ +TGHIM QQA  I+V + +       G+  W       I+  LYW ++SNM  + T
Sbjct: 147 RTIMLTGHIMFQQAGLIAVTLFIF------GYSMWTTIICTAILVSLYWGITSNMMYKPT 200
```

```
Query: 186 QRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPTFLNIFHDTVVASATLML 245
           Q +T G GF+IGHQQQFA W   KVAPF GKKEE++++LKLP +LNIFHD +V++A +M
Sbjct: 201 QEVTDGCGFSIGHQQQFASWIAYKVAPFLGKKEESVEDLKLPGWLNIFHDNIVSTAIVMT 260

Query: 246 VFFGGILAVLGPDIMSNVKLIGPGAFVPTKQAFFMYILQTSLTFSVYLFILMQGVRMFVT 305
           +FFG IL   G D +           +  K  + +YILQT  +F+V +FI+ QGVRMFV
Sbjct: 261 IFFGAILLSFGIDTVQ---------AMAGKVHWTVYILQTGFSFAVAIFIITQGVRMFVA 311

Query: 306 ELTNAFQGISNKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLVVFKNPI 365
           EL+ AF GIS +L+PG+  A+D AA Y F + NAV+ GF +G IGQLI + +LV   + I
Sbjct: 312 ELSEAFNGISQRLIPGAVLAIDCAAIYSF-APNAVVWGFMWGTIGQLIAVGILVACGSSI 370

Query: 366 LIITGFVPVFFDNAAIAVYADKRGGWKAAVALSFISGIIQVALGAVAVGLLGLAGGYHGN 425
           LII GF+P+FF NA I V+A+  GGW+AA+ +  + G+I++     AV L G++  + G
Sbjct: 371 LIIPGFIPMFFSNATIGVFANHFGGWRAALKICLVMGMIEIFGCVWAVKLTGMS-AWMGM 429

Query: 426 IDFEFPWLAFGYIFKYLGIAGYVIVCLFFLA 456
            D+          F  +GIA   ++ +  LA
Sbjct: 430 ADWSILAPPMMQGFFSIGIAFMAVIIVIALA 460
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4579> which encodes the amino acid sequence <SEQ ID 4580>. Analysis of this protein sequence reveals the following:

```
          Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.51   Transmembrane  441 - 457  ( 435 - 465)
   INTEGRAL    Likelihood = -7.80   Transmembrane  344 - 360  ( 339 - 376)
   INTEGRAL    Likelihood = -7.64   Transmembrane  238 - 254  ( 235 - 261)
   INTEGRAL    Likelihood = -5.63   Transmembrane  105 - 121  ( 100 - 127)
   INTEGRAL    Likelihood = -5.52   Transmembrane  138 - 154  ( 137 - 155)
   INTEGRAL    Likelihood = -5.20   Transmembrane  400 - 416  ( 392 - 422)
   INTEGRAL    Likelihood = -4.78   Transmembrane   18 -  34  (  14 -  39)
   INTEGRAL    Likelihood = -2.97   Transmembrane  365 - 381  ( 365 - 383)
   INTEGRAL    Likelihood = -1.49   Transmembrane  160 - 176  ( 160 - 177)
   INTEGRAL    Likelihood = -0.53   Transmembrane   41 -  57  (  41 -  57)

----- Final Results -----
               bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC77150 GB:AE000491 orf, hypothetical protein [Escherichia
           coli]
 Identities = 182/461 (39%), Positives = 279/461 (60%), Gaps = 25/461 (5%)

Query: 1     MEMLLAPLNWFSQNILQNPAFFVGLLVLIGYLLLKKPIYEVFAGFVKATVGYLILNVGAG 60
           ME+L      F   ++ N    +G++  +GY+LL+K +  +  G +K  +G+++L  G+G
Sbjct: 20    MEILYNIFTVFFNQVMTNAPLLLGIVTCLGYILLRKSVSVIIKGTIKTIIGFMLLQAGSG 79

Query: 61    GLVTTFRPILVALAKKFELKAAVIDPYFGLAAANTKLEEMG-FISVATTALLIGFGVNIL 119
            L +TF+P++  +++ + +  A+ D Y  + A    ++ MG   S    A+L+   +NI
Sbjct: 80    ILTSTFKPVVAKMSEVYGINGAISDTYASMMAT---IDRMGDAYSWVGYAVLLALALNIC 136

Query: 120 LVALRKVTKVRTLFITGHIMVQQAATISVFVLLLIPQFQNAFGAWAV----GIICGLYWA 175
            V LR++T +RT+ +TGHIM QQA  I+V + +        +  W       I+  LYW
Sbjct: 137 YVLLRRITGIRTIMLTGHIMFQQAGLIAVTLFIF------GYSMWTTIICTAILVSLYWG 190

Query: 176 ISSNMTVEATQRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPTFLNIFHD 235
           I+SNM  + TQ +T G GF+IGHQQQFA W   KVAPF GKKEE++++LKLP +LNIFHD
Sbjct: 191 ITSNMMYKPTQEVTDGCGFSIGHQQQFASWIAYKVAPFLGKKEESVEDLKLPGWLNIFHD 250

Query: 236 TVVASATLMLVFFGAILAVLGPDIMSDVDLIGPGAFNPAKQAFFMYILQTSLTFSVYLFI 295
            +V++A +M +FFGAIL   G D +  +           K  + +YILQT  +F+V +FI
Sbjct: 251 NIVSTAIVMTIFFGAILLSFGIDTVQAM---------AGKVHWTVYILQTGFSFAVAIFI 301
```

```
Query: 296 LMQGVRMFVSELTNAFQGISSKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITI 355
           + QGVRMFV+EL+ AF GIS +L+PG+  A+D AA Y F + NAV+ GF +G IGQLI +
Sbjct: 302 ITQGVRMFVAELSEAFNGISQRLIPGAVLAIDCAAIYSF-APNAVVWGFMWGTIGQLIAV 360

Query: 356 ALLVIFKNPILIITGFVPVFFDNAAIAVYADKRGGWKAAVALSFISGILQVALGAVAVGL 415
            +LV   + ILII GF+P+FF NA I V+A+  GGW+AA+ +  + G++++     AV L
Sbjct: 361 GILVACGSSILIIPGFIPMFFSNATIGVFANHFGGWRAALKICLVMGMIEIFGCVWAVKL 420

Query: 416 LGLTGGYHGNIDLVLPWLPFGYLFKFLGIAGYVLVCIFLLA 456
            G++  + G  D  +   P    F  +GIA   ++ +  LA
Sbjct: 421 TGMS-AWMGMADWSILAPPMMQGFFSIGIAFMAVIIVIALA 460
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 437/476 (91%), Positives = 457/476 (95%)

Query: 1   MENFLAPLNWFSQNILQNPAFFVGLLVLIGYLLLKKPLHDVFAGFIKATVGYLILNVGAG 60
           ME  LAPLNWFSQNILQNPAFFVGLLVLIGYLLLKKP+++VFAGF+KATVGYLILNVGAG
Sbjct: 1   MEMLLAPLNWFSQNILQNPAFFVGLLVLIGYLLLKKPIYEVFAGFVKATVGYLILNVGAG 60

Query: 61  GLVNTFRPILVALAKKFNLEAAVIDPYFGLASANAKLETMGFISVATTALLIGFGINILL 120
           GLV TFRPILVALAKKF L+AAVIDPYFGLA+AN KLE MGFISVATTALLIGFG+NILL
Sbjct: 61  GLVTTFRPILVALAKKFELKAAVIDPYFGLAAANTKLEEMGFISVATTALLIGFGVNILL 120

Query: 121 VALRKVTKVRTLFITGHIMVQQAATISVFVLLLIPQLRNGFGAWAVGIICGLYWAVSSNM 180
           VALRKVTKVRTLFITGHIMVQQAATISVFVLLLIPQ +N FGAWAVGIICGLYWA+SSNM
Sbjct: 121 VALRKVTKVRTLFITGHIMVQQAATISVFVLLLIPQFQNAFGAWAVGIICGLYWAISSNM 180

Query: 181 TVEATQRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPTFLNIFHDTVVAS 240
           TVEATQRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPTFLNIFHDTVVAS
Sbjct: 181 TVEATQRLTGGGGFAIGHQQQFAIWFVDKVAPFFGKKEENLDNLKLPTFLNIFHDTVVAS 240

Query: 241 ATLMLVFFGGILAVLGPDIMSNVKLIGPGAFVPTKQAFFMYILQTSLTFSVYLFILMQGV 300
           ATLMLVFFG ILAVLGPDIMS+V LIGPGAF P KQAFFMYILQTSLTFSVYLFILMQGV
Sbjct: 241 ATLMLVFFGAILAVLGPDIMSDVDLIGPGAFNPAKQAFFMYILQTSLTFSVYLFILMQGV 300

Query: 301 RMFVTELTNAFQGISNKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLVV 360
           RMFV+ELTNAFQGIS+KLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLV+
Sbjct: 301 RMFVSELTNAFQGISSKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLVI 360

Query: 361 FKNPILIITGFVPVFFDNAAIAVYADKRGGWKAAVALSFISGIIQVALGAVAVGLLGLAG 420
           FKNPILIITGFVPVFFDNAAIAVYADKRGGWKAAVALSFISGI+QVALGAVAVGLLGL G
Sbjct: 361 FKNPILIITGFVPVFFDNAAIAVYADKRGGWKAAVALSFISGILQVALGAVAVGLLGLTG 420

Query: 421 GYHGNIDFEFPWLAFGYIFKYLGIAGYVIVCLFFLAIPQLQFMKSKDKEAYYRGDA 476
           GYHGNID   PWL FGY+FK+LGIAGYV+VC+F LAIPQLQF K+KDKEAYYRG+A
Sbjct: 421 GYHGNIDLVLPWLPFGYLFKFLGIAGYVLVCIFLLAIPQLQFAKAKDKEAYYRGEA 476
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1489

A DNA sequence (GBSx1575) was identified in *S.agalactiae* <SEQ ID 4581> which encodes the amino acid sequence <SEQ ID 4582>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1225(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG34743 GB:AE000033 similar to PTS system: EIIB [Mycoplasma pneumoniae]
 Identities = 40/89 (44%), Positives = 62/89 (68%), Gaps = 1/89 (1%)

Query: 4  VLTACGNGMGSSMVIKMKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNHLIHE 63
          ++ ACGNGMG+SM+IK+KVE  +++LG +    A  S+G+ KG+  + DI+++S HL  E
Sbjct: 8  IIAACGNGMGTSMLIKIKVEKIMKELGYTAKVEA-LSMGQTKGMEHSADIIISSIHLTSE 66

Query: 64 LDGRTKGHLVGLDNLMDDNEIKTKLQEIL 92
           +   K  +VG+ NLMD+NEIK  L ++L
Sbjct: 67 FNPNAKAKIVGVLNLMDENEIKQALSKVL 95
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4583> which encodes the amino acid sequence <SEQ ID 4584>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0977(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 85/92 (92%), Positives = 90/92 (97%)

Query: 1  MVKVLTACGNGMGSSMVIKMKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNHL 60
          MVKVLTACGNGMGSSMVIKMKVENALRQLGV++ +SASCSVGEAKGLA+ YDIVVASNHL
Sbjct: 1  MVKVLTACGNGMGSSMVIKMKVENALRQLGVTDIQSASCSVGEAKGLASGYDIVVASNHL 60

Query: 61 IHELDGRTKGHLVGLDNLMDDNEIKTKLQEIL 92
          IHELDGRTKGHLVGLDNLMDDNEIKTKLQE+L
Sbjct: 61 IHELDGRTKGHLVGLDNLMDDNEIKTKLQEVL 92
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1490

A DNA sequence (GBSx1576) was identified in *S.agalactiae* <SEQ ID 4585> which encodes the amino acid sequence <SEQ ID 4586>. This protein is predicted to be a pentitol phosphotransferase enzyme ii, a component (ptxA). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3309(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC77152 GB:AE000491 putative PTS system enzyme II A component
          [Escherichia coli K12]
 Identities = 64/150 (42%), Positives = 97/150 (64%), Gaps = 2/150 (1%)

Query: 1    MNLKQAFIENDSIRLKLSASDWKEAIKLSIDPLIESGAVDAEYYDAIIESTEEFGPYYIL 60
            M L+ +  EN SIRL+  A  W+EA+K+ +D L+ +  V+  YY AI++  E+FGPY+++
Sbjct: 1    MKLRDSLAENKSIRLQAEAETWQEAVKIGVDLLVAADVVEPRYYQAILDGVEQFGPYFVI 60

Query: 61   MPGMAMPHARPEAGVKRDAFSLITLTEPVVF--PDGKEVSVLLALAATSSAIHTSVAIPQ 118
```

```
            PG+AMPH RPE GVK+  FSL+TL +P+ F   D   V +L+ +AA  +  H  V I Q
Sbjct: 61  APGLAMPHGRPEEGVKKTGFSLVTLKKPLEFNHDDNDPVDILITMAAVDANTHQEVGIMQ 120

Query: 119 IIALFELENSIQRLTECQEAKEVLAMVEES 148
           I+ LFE E +  RL  C+  +EVL +++ +
Sbjct: 121 IVNLFEDEENFDRLRACRTEQEVLDLIDRT 150
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4587> which encodes the amino acid sequence <SEQ ID 4588>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2287(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 113/161 (70%), Positives = 137/161 (84%)

Query: 1    MNLKQAFIENDSIRLKLSASDWKEAIKLSIDPLIESGAVDAEYYDAIIESTEEFGPYYIL 60
           MNLKQAFI+N+SIRL LSA  W+EA++L++ PLI+S AV + YYDAII STE++GPYY+L
Sbjct: 1    MNLKQAFIDNNSIRLGLSADTWQEAVRLAVQPLIDSKAVTSAYYDAIIASTEKYGPYYVL 60

Query: 61   MPGMAMPHARPEAGVKRDAFSLITLTEPVVFPDGKEVSVLLALAATSSAIHTSVAIPQII 120
           MPGMAMPHA    GV R+AF+LITLT+PV F DGKEVSVLL LAAT  +IHT+VAIPQI+
Sbjct: 61   MPGMAMPHAEAGLGVNRNAFALITLTKPVTFSDGKEVSVLLTLAATDPSIHTTVAIPQIV 120

Query: 121 ALFELENSIQRLTECQEAKEVLAMVEESKNSPYLEGLDLES 161
           ALFEL+N+I+RL  CQ  KEVL MVEESK+SPYLEG+DL +
Sbjct: 121 ALFELDNAIERLVACQSPKEVLEMVEESKDSPYLEGMDLNA 161
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1491

A DNA sequence (GBSx1577) was identified in *S.agalactiae* <SEQ ID 4589> which encodes the amino acid sequence <SEQ ID 4590>. This protein is predicted to be probable hexulose-6-phosphate synthase. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1584(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC77153 GB:AE000491 probable hexulose-6-phosphate synthase
          [Escherichia coli K12]
 Identities = 108/217 (49%), Positives = 141/217 (64%), Gaps = 3/217 (1%)

Query: 5    LPNLQVALDHSDLQGAIKAAVSVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPDKIIVAD 64
           LP LQVALD+  +  A +    +  EVD+IE GT+  +  G   V  L++L+P KI++AD
Sbjct: 3    LPMLQVALDNQTMDSAYETTRLIAEEVDIIEVGTILCVGEGVRAVRDLKALYPHKIVLAD 62

Query: 65   TKCADAGGTVAKNNAVRGADWMTCICCATIPTMEAALKAIKEERGDRGEIQIELYGDWTY 124
            K ADAG  +++       ADW+T ICCA I T + AL   KE  GD   +QIEL G WT+
Sbjct: 63   AKIADAGKILSRMCFEANADWVTVICCADINTAKGALDVAKEFNGD---VQIELTGYWTW 119
```

```
Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKKLIDMGFRVSVTGGLSTDTLQL 184
           EQAQQW DAGI Q +YH+SRDA  AG  WGE D+  +K+L DMGF+V+VTGGL+ + L L
Sbjct: 120 EQAQQWRDAGIGQVVYHRSRDAQAAGVAWGEADITAIKRLSDMGFKVTVTGGLALEDLPL 179

Query: 185 FEGVDVFTFIAGRGITEADDPAAAARAFKDEIKRIWG 221
           F+G+ +   FIAGR I +A  P  AAR FK  I  +WG
Sbjct: 180 FKGIPIHVFIAGRSIRDAASPVEAARQFKRSIAELWG 216
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4591> which encodes the amino acid sequence <SEQ ID 4592>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1473(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 206/217 (94%), Positives = 212/217 (96%)

Query: 5   LPNLQVALDHSDLQGAIKAAVSVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPDKIIVAD 64
            +PNLQVALDHSDLQGA+KAAV+VGHEVDVIEAGTVCLLQVGSELVEVLRSLFP+KIIVAD
Sbjct: 4   IPNLQVALDHSDLQGAVKAAVAVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPEKIIVAD 63

Query: 65  TKCADAGGTVAKNNAVRGADWMTCICCATIPTMEAALKAIKEERGDRGEIQIELYGDWTY 124
            TKCADAGGTVAKNNA RGADWMTCICCATIPTMEAALKAIKEERGDRGEIQIELYGDWTY
Sbjct: 64  TKCADAGGTVAKNNAKRGADWMTCICCATIPTMEAALKAIKEERGDRGEIQIELYGDWTY 123

Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKKLIDMGFRVSVTGGLSTDTLQL 184
            EQAQ WLDAGISQAIYHQSRDALLAGETWGEKDLNKVK LIDMGFRVSVTGGL  DTL+L
Sbjct: 124 EQAQLWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKTLIDMGFRVSVTGGLDVDTLRL 183

Query: 185 FEGVDVFTFIAGRGITEADDPAAAARAFKDEIKRIWG 221
            FEGVDVFTFIAGRGITEA+DPAAAARAFKDEIKRIWG
Sbjct: 184 FEGVDVFTFIAGRGITEAEDPAAAARAFKDEIKRIWG 220
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1492

A DNA sequence (GBSx1578) was identified in *S.agalactiae* <SEQ ID 4593> which encodes the amino acid sequence <SEQ ID 4594>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4179(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22686 GB:U32783 hexulose-6-phosphate isomerase, putative
          [Haemophilus influenzae Rd]
 Identities = 143/282 (50%), Positives = 199/282 (69%), Gaps = 3/282 (1%)

Query: 5   IGIYEKATPKHFNWLERLQFAKELGFDFVELSIDESDERLARLEWSKEERLELVKAIFET 64
            IGIYEKA PK+  W ERL  AK  GF+F+E+SIDES++RL+RL W+K ER+ L ++I ++
Sbjct: 6   IGIYEKALPKNITWQERLSLAKACGFEFIEMSIDESNDRLSRLNWTKSERIALHQSIIQS 65
```

```
Query: 65  GVRVPTITFSGHRRFPMGSNNPEKEARAMDMMKKCIVFAQDIGIRNIQLAGYDVYYEEKS 124
           G+ +P++  S HRRFP GS + +   ++ ++M+K I  + ++GIR IQLAGYDVYYE++
Sbjct: 66  GITIPSMCLSAHRRFPFGSKDKKIRQKSFEIMEKAIDLSVNLGIRTIQLAGYDVYYEKQD 125

Query: 125 PETRARFIKNLRQACTWAEEAQVILSIEIMDDPFMNSIEKYLAVEKEIDSPYLFVYPDTG 184
            ET   F + +  A T A  AQV L++EIMD PFM+SI ++   +  I+SP+  VYPD G
Sbjct: 126 EETIKYFQEGIEFAVTLAASAQVTLAVEIMDTPFMSSISRWKKWDTIINSPWFTVYPDIG 185

Query: 185 NVSAWHNDLWSEFYNGHRSIAALHIKDTYAVTETSKGQFRDVPFGQGCVDWEEMFAVIKK 244
           N+SAW+N++  E   G   I+A+H+KDTY VTETSKGQFRDVPFGQGCVD+   F+++KK
Sbjct: 186 NLSAWNNNIEEELTLGIDKISAIHLKDTYPVTETSKGQFRDVPFGQGCVDFVHFFSLLKK 245

Query: 245 TNYNGPFLIEMWSENCETVEETRAAIKEAQDFLYPLMEKTGV 286
            NY G FLIEMW+E     EE    I +A+ ++   MEK G+
Sbjct: 246 LNYRGAFLIEMWTEK---NEEPLLEIIQARKWIVQQMEKAGL 284
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4595> which encodes the amino acid sequence <SEQ ID 4596>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1489(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 240/286 (83%), Positives = 271/286 (93%)

Query: 1   MTRPIGIYEKATPKHFNWLERLQFAKELGFDFVELSIDESDERLARLEWSKEERLELVKA 60
           M RPIGIYEKATPK F W ERLQFAK+LGFDFVE+S+DESD RLARLEW+KEERL+LVKA
Sbjct: 15  MARPIGIYEKATPKQFTWRERLQFAKDLGFDFVEMSVDESDARLARLEWTKEERLDLVKA 74

Query: 61  IFETGVRVPTITFSGHRRFPMGSNNPEKEARAMDMMKKCIVFAQDIGIRNIQLAGYDVYY 120
           I+ETG+R+PTI FSGHRR+P+GSN+P  EA+++ +MK+CI  AQD+G+R IQLAGYDVYY
Sbjct: 75  IYETGIRIPTICFSGHRRYPLGSNDPAIEAKSLKLMKQCIELAQDLGVRTIQLAGYDVYY 134

Query: 121 EEKSPETRARFIKNLRQACTWAEEAQVILSIEIMDDPFMNSIEKYLAVEKEIDSPYLFVY 180
           E+KSPETRARFIKNLRQ+C WAEEAQV+LSIEIMDDPF+NSIEKYLAVEKEIDSPYLFVY
Sbjct: 135 EKKSPETRARFIKNLRQSCDWAEEAQVMLSIEIMDDPFINSIEKYLAVEKEIDSPYLFVY 194

Query: 181 PDTGNVSAWHNDLWSEFYNGHRSIAALHIKDTYAVTETSKGQFRDVPFGQGCVDWEEMFA 240
           PD GNVSAWHNDLWSEFYNGH+SIAALH+KDTYAVTETSKGQFRDVPFGQGCVDW+E+FA
Sbjct: 195 PDAGNVSAWHNDLWSEFYNGHKSIAALHLKDTYAVTETSKGQFRDVPFGQGCVDWQELFA 254

Query: 241 VIKKTNYNGPFLIEMWSENCETVEETRAAIKEAQDFLYPLMEKTGV 286
           V+KKTNYNGPFLIEMWSENC+TVEET+AAIKEAQDFLYPL+EK G+
Sbjct: 255 VLKKTNYNGPFLIEMWSENCDTVEETKAAIKEAQDFLYPLIEKAGL 300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1493

A DNA sequence (GBSx1579) was identified in *S.agalactiae* <SEQ ID 4597> which encodes the amino acid sequence <SEQ ID 4598>. This protein is predicted to be L-ribulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2559(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase
          [Bacillus stearothermophilus]
 Identities = 143/229 (62%), Positives = 176/229 (76%), Gaps = 2/229 (0%)

Query: 5   LQEMRERVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDL 64
           L+E+++ V EAN  LP + LV FTWGNVS +DRE GL+VIKPSGV YD+LT ++MVV DL
Sbjct: 2   LEELKQAVLEANLQLPQYRLVTFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDL 61

Query: 65  EGNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHTHSTEAVGWAQAGRDIPFYGTTHADYF 124
            GN+VEGDL PSSD PTH+ LYK +P +GGIVHTHST A  WAQAG+ IP  GTTHADYF
Sbjct: 62  TGNVVEGDLKPSSDTPTHLWLYKQFPGIGGIVHTHSTWATVWAQAGKGIPALGTTHADYF 121

Query: 125 YGPVPCARSLSEDEVNTAYEKETGSVIIEEFERRDLDPMAVPGIVVRNHGPFTWGKDPAQ 184
           YG +PC R ++ +E+  AYE ETG VI E F  R LDP+ +PG++V  HGPF WGKDPA
Sbjct: 122 YGEIPCTRPMTNEEIQGAYELETGKVITETF--RFLDPLQMPGVLVHGHGPFAWGKDPAN 179

Query: 185 AVYHSVVLEEVAKMNRFTEQINPRVEPAPKYIMDKHYLRKHGPNAYYGQ 233
           AV+++VVLEEVAKM   T  +NP   +P  + ++D+HYLRKHG NAYYGQ
Sbjct: 180 AVHNAVVLEEVAKMAARTYMLNPNAKPISQTLLDRHYLRKHGANAYYGQ 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4599> which encodes the amino acid sequence <SEQ ID 4600>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2257(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 207/234 (88%), Positives = 220/234 (93%)

Query: 1   MAKSLQEMRERVCEANKSLPVHSLVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMV 60
           MAK+LQEMRERVC ANKSLP H LVKFTWGNVSEV RE G IVIKPSGVDYD LTPENMV
Sbjct: 1   MAKNLQEMRERVCAANKSLPQHGLVKFTWGNVSEVCRELGRIVIKPSGVDYDLLTPENMV 60

Query: 61  VTDLEGNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHTHSTEAVGWAQAGRDIPFYGTTH 120
           VTDL+GN+VEGDLNPSSDLPTHV+LYKAWPEVGGIVHTHSTEAVGWAQAGRDIPFYGTTH
Sbjct: 61  VTDLDGNVVEGDLNPSSDLPTHVELYKAWPEVGGIVHTHSTEAVGWAQAGRDIPFYGTTH 120

Query: 121 ADYFYGPVPCARSLSEDEVNTAYEKETGSVIIEEFERRDLDPMAVPGIVVRNHGPFTWGK 180
           ADYFYGPVPCARSL++ EV+ AYE+ETG+VI+EEF +R LDPMAVPGIVVRNHGPFTWGK
Sbjct: 121 ADYFYGPVPCARSLTKAEVDGAYEQETGNVILEEFSKRGLDPMAVPGIVVRNHGPFTWGK 180

Query: 181 DPAQAVYHSVVLEEVAKMNRFTEQINPRVEPAPKYIMDKHYLRKHGPNAYYGQK 234
            P QAVYHSVVLEEVA+MNR TEQINPRVEPAP+YIMDKHYLRKHGPNAYYGQK
Sbjct: 181 TPEQAVYHSVVLEEVARMNRLTEQINPRVEPAPRYIMDKHYLRKHGPNAYYGQK 234
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1494

A DNA sequence (GBSx1580) was identified in *S.agalactiae* <SEQ ID 4601> which encodes the amino acid sequence <SEQ ID 4602>. This protein is predicted to be transaldolase (tal). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4232(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10149> which encodes amino acid sequence <SEQ ID 10150> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB98962 GB:U67539 transaldolase [Methanococcus jannaschii]
 Identities = 124/214 (57%), Positives = 157/214 (72%)

Query: 19  MKYFLDTADVSEIRRLNRLGIVDGVTTNPTIISREGRDFKEVINEICQIVDGPVSAEVTG 78
            MK+FLDTA+V EI++   LG+VDGVTTNPT++++EGRDF EV+ EIC+IV+GPVSAEV
Sbjct: 1   MKFFLDTANVEEIKKYAELGLVDGVTTNPTLVAKEGRDFYEVVKEICEIVEGPVSAEVIS 60

Query: 79  LTCDEMVTEAREIAKWSPNVVVKIPMTEEGLAAVSQLSKEGIKTNVTLIFTVAQGLSAMK 138
               + MV EARE+AK + N+V+KIPMT++G+ AV  LS EGIKTNVTL+F+  Q L A K
Sbjct: 61  TDAEGMVKEARELAKLADNIVIKIPMTKDGMKAVKILSAEGIKTNVTLVFSPLQALVAAK 120

Query: 139 AGATFISPFVGRLEDIGTDAYALIRDLRHIIDFYGFQSEIIAASIRGLAHVEGVAKCGAH 198
            AGAT++SPFVGRL+DIG     LI D+  I   Y +++E+I AS+R   HV   AK GA
Sbjct: 121 AGATYVSPFVGRLDDIGHVGMKLIEDVVKIYKNYDIKTEVIVASVRHPWHVLEAAKIGAD 180

Query: 199 IATIPDKTFASLFTHPLTDKGIETFLKDWDSFKK 232
            IAT+P      LF HPLTD G+E FLKDWD + K
Sbjct: 181 IATMPPAVMDKLFNHPLTDIGLERFLKDWDEYLK 214
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4603> which encodes the amino acid sequence <SEQ ID 4604>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1902(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 162/214 (75%), Positives = 180/214 (83%)

Query: 19  MKYFLDTADVSEIRRLNRLGIVDGVTTNPTIISREGRDFKEVINEICQIVDGPVSAEVTG 78
            MK+FLDTA+V+ I+ +N LG+VDGVTTNPTIISREGRDF+ VI EIC IVDGP+SAEVTG
Sbjct: 1   MKFFLDTANVAAIKAINELGVVDGVTTNPTIISREGRDFETVIKEICDIVDGPISAEVTG 60

Query: 79  LTCDEMVTEAREIAKWSPNVVVKIPMTEEGLAAVSQLSKEGIKTNVTLIFTVAQGLSAMK 138
            LT D MV EAR IAKW  NVVVKIPMT EGL A + LSKEGIKTNVTLIFTV+QGL AMK
Sbjct: 61  LTADAMVEEARSIAKWHDNVVVKIPMTTEGLKATNILSKEGIKTNVTLIFTVSQGLMAMK 120

Query: 139 AGATFISPFVGRLEDIGTDAYALIRDLRHIIDFYGFQSEIIAASIRGLAHVEGVAKCGAH 198
            AGAT+ISPF+GRLEDIGTDAY LI DLR IID Y FQ+EIIAASIR  AHVE VAK GAH
Sbjct: 121 AGATYISPFIGRLEDIGTDAYQLISDLREIIDLYDFQAEIIAASIRTTAHVEAVAKLGAH 180
```

```
Query: 199 IATIPDKTFASLFTHPLTDKGIETFLKDWDSFKK 232
           IATIPD  FA +  HPLT  G++TF++DW SFKK
Sbjct: 181 IATIPDPLFAKMTQHPLTTNGLKTFMEDWASFKK 214
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1495

A DNA sequence (GBSx1581) was identified in *S.agalactiae* <SEQ ID 4605> which encodes the amino acid sequence <SEQ ID 4606>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1263(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14129 GB:Z99115 transcriptional regulator (LacI family)
           [Bacillus subtilis]
 Identities = 108/331 (32%), Positives = 188/331 (56%), Gaps = 12/331 (3%)

Query: 6   TISDIANLVGVSKATVSYYLNGNYKKMSLQTKEKIRLAIKETGYQPSKIAQSLVTKNTRT 65
           TI D+A   GVSK+TVS Y+NG    +S +  + I+ AI E  Y+PSK+AQ L  K ++
Sbjct: 10  TIKDVAECAGVSKSTVSRYINGKIDAISPEKVKNIKKAIAELNYRPSKMAQGLKIKKSKL 69

Query: 66  IGVVIADITNPFISSVMKGIHDTCQQFGYSVNFTNSDNDIDIELENLNRLNQQNVSGIIL 125
           IG V+ADITNPF  +  +G+ + C Q+GYS+   N+DN  + E E L +L   +V G+IL
Sbjct: 70  IGFVVADITNPFSVAAFRGVEEVCDQYGYSIMVCNTDNSPEKEREMLLKLEAHSVEGLIL 129

Query: 126 DSVDPNHSFIETLSNDRL--VMVDRQAKDIKVDTVASDNKESTQIFLEKMQEAGYHDIYF 183
           ++   N   +   ++  +++DR+  D+K+DTV +DN+  T+  L+K+   GY D+
Sbjct: 130 NATGENKDVLRAFAEQQIPTILIDRKLPDLKLDTVTTDNRWITKEILQKVYSKGYTDVAL 189

Query: 184 VTYPIEGISTRELRYEGFKEVVS-SNPDKLIIITE-DGSTQRILDI------IEHSEQKP 235
            T PI  IS R  R   ++E+ S  N + L+ + E D   +  L         E  EQK
Sbjct: 190 FTEPISSISPRAERAAVYQEMASVQNVNGLVRLHEIDVKDKEQLKAELRSFHKEMPEQKK 249

Query: 236 GFLMMNGPTLLNFMKKLNQSTVSYPEDYGLGSYEDLEWMQVLTPNVSCIKQDSYGIGCLA 295
             L +NG  +L  +  + +  +  P+D G+  ++D EW +++ P ++ I Q S+ +G  A
Sbjct: 250 AILALNGLIMLKIISCMEELGLRIPQDIGIAGFDDTEWYKLIGPGITTIAQPSHDMGRTA 309

Query: 296 AQCLIEKISQGNEPTTARLLEVKNQIVIRQS 326
            + ++++I    +    + +E++ ++++R+S
Sbjct: 310 MERVLKRIE--GDKGAPQTIELEAKVIMRKS 338
```

There is also homology to SEQ ID 2366.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1496

A DNA sequence (GBSx1582) was identified in *S.agalactiae* <SEQ ID 4607> which encodes the amino acid sequence <SEQ ID 4608>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
 bacterial cytoplasm --- Certainty=0.1661(Affirmative) < succ>
  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1497

A DNA sequence (GBSx1583) was identified in *S.agalactiae* <SEQ ID 4609> which encodes the amino acid sequence <SEQ ID 4610>. This protein is predicted to be GLYCERATE DEHYDROGENASE. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB50351 GB:AJ248287 GLYCERATE DEHYDROGENASE [Pyrococcus abyssi]
 Identities = 123/325 (37%), Positives = 192/325 (58%), Gaps = 8/325 (2%)

Query: 1   MDKKKILVTGIVPKEGLRKLMDRFDVTYSED-RPFSRDYVLEHLSEYDGWLLM-GQKGDK 58
           M K ++ +T  +P+ G+  L   F+V   ED R   R+ +LE + + D  + M  ++ D+
Sbjct: 1   MSKPRVFITREIPEVGIEMLEKEFEVEVWEDEREIPREILLEKVKDVDALVTMLSERIDR 60

Query: 59  EMIDAGENLQIISLNAVGFDHVDTAYAKEKGIIVSNSPQAVRVPTAEMTFALILAASKRL 118
           E+ +    L+I++  AVG+D++D   A ++GI V+N+P  +   TA++ FAL+LA ++ L
Sbjct: 61  EVFERAPRLRIVANYAVGYDNIDVEEATKRGIYVTNTPGVLTDATADLAFALLLATARHL 120

Query: 119 AFYDSIVRSGEW----IDPSEQRYQGLTLQGSTLGIYGMGRIGLTVANFAKAFGMTVVYN 174
              D   RSGEW    +    + + G  + G T+GI G GRIG  +A  A+ F M ++Y
Sbjct: 121 VKGDKFTRSGEWKKRGVAWHPKWFLGYDVYGKTIGIIGFGRIGQAIAKRARGFDMRILYY 180

Query: 175 DVYRLPEDKEKELGVTYLEFDQLIKTADVITIHAPALPSTIHKFNKDVFAKMKNRSYLIN 234
              R PE  EKEL    +   D+L++ +D + +  P    T H  N++    MK  + LIN
Sbjct: 181 SRTRKPE-VEKELNAEFKPLDELLRESDFVVLAVPLNKETYHMINEERLKMMKRTAILIN 239

Query: 235 AARGPIVSEEALIEALKEGEIAGAGLDVFENEPQVSEGLRSLDNVIMSPHAGTGTIEGRR 294
            ARG ++  +ALI+ALKEG IAGAGLDV+E EP  +E L SLDNV+++PH G+ T   R
Sbjct: 240 VARGKVIDTKALIKALKEGWIAGAGLDVYEEEPYYNEELFSLDNVVLTPHIGSATFGARE 299

Query: 295 TLAEEAADNIIAFFDGK-PQNIVNK 318
            +A+  A+N+IAF  G+ P  +VN+
Sbjct: 300 GMAKLVAENLIAFKRGEVPPTLVNR 324
```

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1498

A DNA sequence (GBSx1585) was identified in *S.agalactiae* <SEQ ID 4611> which encodes the amino acid sequence <SEQ ID 4612>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1898(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1499

A DNA sequence (GBSx1586) was identified in *S.agalactiae* <SEQ ID 4613> which encodes the amino acid sequence <SEQ ID 4614>. This protein is predicted to be PTS system, galactitol specific IIC component. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.27   Transmembrane  254 - 270 ( 245 - 277)
   INTEGRAL    Likelihood = -9.24   Transmembrane   77 -  93 (  71 - 100)
   INTEGRAL    Likelihood = -9.24   Transmembrane  367 - 383 ( 364 - 386)
   INTEGRAL    Likelihood = -8.28   Transmembrane   32 -  48 (  26 -  54)
   INTEGRAL    Likelihood = -7.38   Transmembrane  186 - 202 ( 182 - 215)
   INTEGRAL    Likelihood = -6.26   Transmembrane  158 - 174 ( 151 - 180)
   INTEGRAL    Likelihood = -5.79   Transmembrane  279 - 295 ( 276 - 296)
   INTEGRAL    Likelihood = -1.12   Transmembrane  342 - 358 ( 342 - 359)
   INTEGRAL    Likelihood = -0.00   Transmembrane  308 - 324 ( 308 - 324)

----- Final Results -----
               bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8825> which encodes amino acid sequence <SEQ ID 8826> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      8.30
GvH: Signal Score (-7.5): 2.97
     Possible site: 58
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 9 value: -13.27 threshold:  0.0
   INTEGRAL    Likelihood =-13.27   Transmembrane  321 - 337 ( 312 - 344)
   INTEGRAL    Likelihood = -9.24   Transmembrane  144 - 160 ( 138 - 167)
   INTEGRAL    Likelihood = -9.24   Transmembrane  434 - 450 ( 431 - 453)
   INTEGRAL    Likelihood = -8.28   Transmembrane   99 - 115 (  93 - 121)
   INTEGRAL    Likelihood = -7.38   Transmembrane  253 - 269 ( 249 - 282)
   INTEGRAL    Likelihood = -6.26   Transmembrane  225 - 241 ( 218 - 247)
   INTEGRAL    Likelihood = -5.79   Transmembrane  346 - 362 ( 343 - 363)
   INTEGRAL    Likelihood = -1.12   Transmembrane  409 - 425 ( 409 - 426)
   INTEGRAL    Likelihood = -0.00   Transmembrane  375 - 391 ( 375 - 391)
   PERIPHERAL  Likelihood =  0.69     188
 modified ALOM score:   3.15
```

```
*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03909 GB:AP001507 PTS system, galactitol-specific enzyme II,
          C component [Bacillus halodurans]
 Identities = 92/347 (26%), Positives = 173/347 (49%), Gaps = 15/347 (4%)

Query: 1   MVKTTGLHLPIVDIGWQAGSLTAFSSEIGLSFFVFGLLIELGLFLLGITRVFVPSNLWNN 60
           MV   G+ L ++D+GW A S  A++S +       GL++ + + +   T+  +  ++WN
Sbjct: 70  MVDRLGVDLNVIDVGWPATSSIAWASVVAAFIIPLGLIVNVIMLVTKTTKT-MNVDIWNF 128

Query: 61  FGYMIWGTMAYAATGNFILSFAFMVFVLLYSLVMSEVLADRWSEYYGVKNATINSIHNIE 120
           + Y     + Y  + + I +    V   + +L +++  A   SE+Y +   +I +   I
Sbjct: 129 WHYTFMAAVVYTVSDSIIQALIAAVMFQIVALKVADWTAPMVSEFYELPGVSIATGSTIS 188

Query: 121 TLIPALILDPLWNLLGVNKVKLNPESLKTKLGIFGEPMTLGFILGVIIGVLGSLRNLASI 180
                 ++  +  + G+     +P++++ + GIFGE + +G ILG  IG+L
Sbjct: 189 YAPGIWLVKGIQKIPGIKHWNADPDTIQRRFGIFGESIFIGLILGAAIGLLAGYNV---- 244

Query: 181 DTWGGILGFAVALAAVMTIFPLITGVFASAFAPLAEAVERNKKKESQAEQGALDKKRWFI 240
              G ++   +A+AAVM + P +  +      P++E+      K           +   I
Sbjct: 245 ---GEVIEIGMAMAAVMVLMPRMVKILMEGLMPVSESAREWLNKR-------FGDREIHI 294

Query: 241 AVDDGVGFGEPATIIAGLILVPIMVVISLILPGNEALPVVDLIAIPFMIEAMIAVSKGNI 300
            +D  V  G P+ I   LILVP+ V++++ILPGN  LP  DL  IPF++  ++  ++GNI
Sbjct: 295 GLDAAVLLGHPSVISTALILVPLTVLLAVILPGNALLPFGDLATIPFIVAFIVGAARGNI 354

Query: 301 LKAILNGIIWFSLGLYAASALGPIYTEAVKHFGTALPAGVTLIMSFN 347
           + ++L G I  +L LY A+ + P++T+  ++    +P G  LI S +
Sbjct: 355 IHSVLAGAIMIALSLYMATDIAPVFTKMAENSNFNMPEGSALISSID 401
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1500

A DNA sequence (GBSx1587) was identified in *S.agalactiae* <SEQ ID 4615> which encodes the amino acid sequence <SEQ ID 4616>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1013(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1501

A DNA sequence (GBSx1588) was identified in *S.agalactiae* <SEQ ID 4617> which encodes the amino acid sequence <SEQ ID 4618>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1294(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10147> which encodes amino acid sequence <SEQ ID 10148> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC76604 GB:AE000435 L-xylulose kinase, cryptic [Escherichia
           coli K12]
 Identities = 156/496 (31%), Positives = 261/496 (52%), Gaps = 18/496 (3%)

Query: 16  YYLSIDYGGTNTKALIFDKLGHQIAVSSFETLKNETQSGHRQVNLVKTWNAITSAIREVI 75
           Y+L +D GG+  KA ++D+ G +  V          Q G  + ++ + W    + IR ++
Sbjct: 4   YWLGLDCGGSWLKAGLYDREGREAGVQRLPLCALSPQPGWAERDMAELWQCCMAVIRALL 63

Query: 76  QISKLSPEQISAVACIGHGKGLYLLDNKLEPLEQGILSTDNRAKDLAQYFESK--LDNIW 133
             S +S EQI  +     GKGL+LLD   +PL   ILS+D RA ++ + ++     + ++
Sbjct: 64  THSGVSGEQIVGIGISAQGKGLFLLDKNDKPLGNAILSSDRRAMEIVRRWQEDGIPEKLY 123

Query: 134 ELTRQHIFPSQSPVILRWLKDYQPETYKSIGAVLSAKDFIRYKLTGKVQQEYGDASGNHW 193
            LTRQ ++      +LRWLK+++PE Y  IG V+   D++R+ LTG    E  + S ++
Sbjct: 124 PLTRQTLWTGHPVSLLRWLKEHEPERYAQIGCVMMTHDYLRWCLTGVKGCEESNISESNL 183

Query: 194 INFQTGTYDPAILDFFGIREIENSLPELIDSADLVPGGISSQAAKETGLVEGTPVVGGLF 253
            N   G YDP + D+ GI EI ++LP ++ SA++  G I++Q A  TGL  GTPVVGGLF
Sbjct: 184 YNMSLGEYDPCLTDWLGIAEINHALPPVVGSAEIC-GEITAQTAALTGLKAGTPVVGGLF 242

Query: 254 DIDACALGSGVLESDTFSVISGTWNINT--YPSLKPAKQDSGLMTSYFPDRRYLLEASSP 311
           D+ + AL +G+ +  T + + GTW + +     L+  +    +   Y  D  +++  +SP
Sbjct: 243 DVVSTALCAGIEDEFTLNAVMGTWAVTSGITRGLRDGEAHPYVYGRYVNDGEFIVHEASP 302

Query: 312 TSAGNLNFMLKMLMHQEIDNAKSSGGSIYDNLEEFLTHTDATHHGLIFFPFLYGSNTSQD 371
           TS+GNL +                G   +D + + +         L F PFLYGSN   +
Sbjct: 303 TSSGNLEWF-----------TAQWGEISFDEINQAVASLPKAGGDLFFLPFLYGSNAGLE 351

Query: 372 ASACFFGLTTKSTKSQMIRAVYEGIAFAHKQHITDLIKSRGSVPKIIRFSGGATNSPAWM 431
            ++ F+G+    T++ +++A+YEG+ F+H  H+ + ++ R +    +R +GG  +S  WM
Sbjct: 352 MTSGFYGMQAIHTRAHLLQAIYEGVVFSHMTHL-NRMRERFTDVHTLRVTGGPAHSDVWM 410

Query: 432 QMFSDILNFPIETVEGTELGGLGGAILARHALDKI-SLKEAVQDMVRVKAIYKPQLSEVK 490
           QM +D+    IE  +  E G  G A+ AR       +  EA +D+        P ++  +
Sbjct: 411 QMLADVSGLRIELPQVEETGCFGAALAARVGTGVYHNFSEAQRDLRHPVRTLLPDMTAHQ 470

Query: 491 GYKKKYHAYQKLLETL 506
            Y+KKY  YQ L+  L
Sbjct: 471 LYQKKYQRYQHLIAAL 486
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1502

A DNA sequence (GBSx1589) was identified in *S.agalactiae* <SEQ ID 4619> which encodes the amino acid sequence <SEQ ID 4620>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG05648 GB:AE004652 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 59/235 (25%), Positives = 104/235 (44%), Gaps = 9/235 (3%)

Query: 23  QVQLIKLVKDLGFSRFEIRQELLQDPDRELPALKAEADFYDINLYYSANEDLIK-GGKVN 81
           Q   + L+   G  R E+R+EL   P  +  AL A      +   +S+  +L +  G++N
Sbjct: 23  QASFLPLLAMAGAQRVELREELFAGPP-DTEALTAAIQLQGLECVFSSPLELWREDGQLN 81

Query: 82  PYLNKGLKEASQLGAPFIKLNVGQTRNLSKEELEPLKEILKSQTIGIKVENNQDPKAATV 141
           P L   L+ A   GA ++K+++G      + +L  L   L    + + VEN+Q P+   +
Sbjct: 82  PELEPTLRRAEACGAGWLKVSLGLLPE--QPDLAALGRRLARHGLQLLVENDQTPQGGRI 139

Query: 142 ENCQYFMTLVKELQIPISFVFDTANWAFINQDLYQAVNNLACDTTYLHCKNFIQVAGKPH 201
           E  + F  L +  Q+ ++  FD  NW +  Q   +A   L     Y+HCK  I+
Sbjct: 140 EVLERFFRLAERQQLDLAMTFDIGNWRWQEQAADEAALRLGRYVGYVHCKAVIRNRDGKL 199

Query: 202 LSKSLFEGEINLTD-LLKSFSNCEYLALEYPTE----LEILKRDVQRLISISNSQ 251
           ++      ++     LL+ F      A+EYP +    L + +R +  L  +   Q
Sbjct: 200 VAVPPSAADLQYWQRLLQHFPEGVARAIEYPLQGDDLLSLSRRHIAALARLGQPQ 254
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1503

A DNA sequence (GBSx1590) was identified in *S.agalactiae* <SEQ ID 4621> which encodes the amino acid sequence <SEQ ID 4622>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0430(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB03939 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 136/511 (26%), Positives = 234/511 (45%), Gaps = 29/511 (5%)

Query: 4   LDKKSYDLLFYLLKLEEPETVMAIANALNQSRRKVYYHLEKINDALPSDVPQIVSYPRV- 62
           LD++S  +L  LL       +  +   LN SRR VY  LEKIN  L     + V   R
Sbjct: 3   LDQRSTFILTQLLHARSYLPIQELTQKLNVSRRTVYNDLEKINSWLEEQGLKAVYKVRSQ 62

Query: 63  GILLTEKQKAACRLLLDEVTDYSYVMKSSERLQLSLVSIVVAKDRVTIDRLMQLNDVSRN 122
           G++L E+ K      L  +  + Y   + ER    ++ ++   + + ++ LM    VSRN
Sbjct: 63  GLILDERAKEEIPTKLRSLKSWHYEYSAQERKAWVVIYLLTRLEPLFLEHLMDRTGVSRN 122
```

```
Query: 123 TILNDLNELRSELAEKEYNLQLQSTKCRGYFLDGHPL----SIIQYLYKLLDDIYHNGSS 178
           T ++D+  L+ EL    ++L L+  +  GY + G       +++ YL + L
Sbjct: 123 TTIDDIKCLKDEL--NNFHLALEFERKDGYTISGDETDKRKALVYYLSQALPQQNWETEL 180

Query: 179 SFIDLFNHKLSQAFGASTYFSKEVLDYFHHYLFISQRSLGKKINSQDGQFMIQILPFILM 238
           S I +F   L         F+ E L   +  +  S++ L  KI   D       L F+L
Sbjct: 181 SPIRIF---LRTKRDNGRIFTIEELQKVYDVISESEKVL--KIQYTDDVLHSLSLRFLLF 235

Query: 239 AYRK-----MRLSPEVQTSLNSDFSLVWQRKEYEIAKELADELEENFQLSLDEIEVGLVA 293
             R      +++ P  +  L          KEYE AK ++ +LE+ F +   + EV  +
Sbjct: 236 MKRVAKGKFIKVHPLEKQVLKGT-------KEYEAAKVMSFKLEQAFGVHYPDEEVLYLT 288

Query: 294 MLMLSFRKDRDN-HLESQ-DYDDMRATLTSFLKELEERYHLHFVHKKDLLRQLLTHCKAL 351
             +LS + +  N  +ES+ +  ++   +TS + + ++   + F  K+ L + L  H K
Sbjct: 289 THILSSKINYANGEIESRKESQELTHIVTSMVNDFQKYACVVFEEKELLEKNLFFHIKPA 348

Query: 352 LYRKRYGIFSVNPLTEHIKDKYEELFAITSSSVKLLEKAWQIKLTDDDVAYLTIHLGGEL 411
            YR +YG+   N + E IK  Y ELF +T   V  LE+     + D++VA++T+H  G +
Sbjct: 349 FYRIKYGLEVENNIAESIKTSYPELFLLTRKVVHYLERYVGKSVNDNEVAFITMHFVGWM 408

Query: 412 RNSQQSPNK-LKLVIVSDEGIAIQKLLLKQCQRYLTNSDIEAVFTTEQYQSVSDLMHVDM 470
           R     P K  K +IV   G+   + L  Q +      DI    +  +Y+     + VD
Sbjct: 409 RREGTIPTKRKKALIVCANGVGTSQFLKNQLEGLFPAVDIIKTCSIREYEKTP--VEVDF 466

Query: 471 VVSTSDALESRFPMLVVHPVLTDDDIIRLIR 501
           ++ST+   E   P+ +V+P+LT+ +  RL++
Sbjct: 467 IISTTSIPEKNVPIFIVNPILTETEKERLLK 497
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4623> which encodes the amino acid sequence <SEQ ID 4624>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0745(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 368/548 (67%), Positives = 456/548 (83%)

Query: 1   MIILDKKSYDLLFYLLKLEEPETVMAIANALNQSRRKVYYHLEKINDALPSDVPQIVSYP 60
           M+ILDKKSYDLL YLLKLE PETVMAI++ALNQSRRKVYY L+KIN ALP  V QI+SYP
Sbjct: 1   MMILDKKSYDLLSYLLKLETPETVMAISHALNQSRRKVYYQLDKINQALPKGVDQIISYP 60

Query: 61  RVGILLTEKQKAACRLLLDEVTDYSYVMKSSERLQLSLVSIVVAKDRVTIDRLMQLNDVS 120
           R+GILLT  QKAACRLLL+EVTDY+YVMKS ER +LS + I V+ +RVTID+LMQ+NDVS
Sbjct: 61  RLGILLTADQKAACRLLLEEVTDYNYVMKSDERRRLSSIYIAVSTERVTIDKLMQINDVS 120

Query: 121 RNTILNDLNELRSELAEKEYNLQLQSTKCRGYFLDGHPLSIIQYLYKLLDDIYHNGSSSF 180
           RNTILNDL ELR EL +K+Y +QL +TK RGY+   HP+++IQYLYKLL D+Y  G++SF
Sbjct: 121 RNTILNDLTELRREELEDKQYKIQLHATKARGYYFGCHPMALIQYLYKLLVDVYQGGNTSF 180

Query: 181 IDLFNHKLSQAFGASTYFSKEVLDYFHHYLFISQRSLGKKINSQDGQFMIQILPFILMAY 240
           ID+FN KLS+  G S YFSK++L YFH YLF+SQ SLGK IN+QD QFM+QILPF+L++Y
Sbjct: 181 IDIFNRKLSEIQGLSVYFSKDILTYFHEYLFLSQASLGKTINTQDSQFMLQILPFMLLSY 240

Query: 241 RKMRLSPEVQTSLNSDFSLVWQRKEYEIAKELADELEENFQLSLDEIEVGLVAMLMLSFR 300
           R MRL  E +++L  +F L+W+RKEY IA++LA EL  NF+L LD+IEV +VAMLMLSFR
Sbjct: 241 RNMRLDSETKSALKQEFHLIWKRKEYHIAQDLARELYHNFKLHLDDIEVSMVAMLMLSFR 300

Query: 301 KDRDNHLESQDYDDMRATLTSFLKELEERYHLHFVHKKDLLRQLLTHCKALLYRKRYGIF 360
           KD+D+H+ESQDYDDMRAT++ F+ +LE RY LHF HK+DLL++L THCKAL+YRK YGIF
Sbjct: 301 KDQDHHVESQDYDDMRATISHFIDQLESRYQLHFTHKQDLLKRLTTHCKALVYRKAYGIF 360
```

```
Query: 361 SVNPLTEHIKDKYEELFAITSSSVKLLEKAWQIKLTDDDVAYLTIHLGGELRNSQQSPNK 420
            VNPLT+H+K+KYEELFA+T S   +LE+ W I LTDDD+AYLTIHLGGELR++     K
Sbjct: 361 LVNPLTDHVKEKYEELFAMTQSCATILEQDWTISLTDDDIAYLTIHLGGELRHNNTEQEK 420

Query: 421 LKLVIVSDEGIAIQKLLLKQCQRYLTNSDIEAVFTTEQYQSVSDLMHVDMVVSTSDALES 480
            KLVIVSD+GI IQKLL KQCQRYL N  IEAVFTTEQYQSV DL+ VDM+V+T+D L++
Sbjct: 421 TKLVIVSDDGIGIQKLLFKQCQRYLANGQIEAVFTTEQYQSVYDLLAVDMIVATTDTLKT 480

Query: 481 RFPMLVVHPVLTDDDIIRLIRFSKKGNCANSNQFTNELEKTIAQYVKEDSERYVLKSKIE 540
           + PML+V+P+L+DDDII+LIRFSK+G  +  ++F+ EL K I   VK++S+RY L SKIE
Sbjct: 481 KIPMLIVNPILSDDDIIKLIRFSKQGRLSEHSRFSTELTKAIEAVVKDESDRYALVSKIE 540

Query: 541 KLIHQELL 548
           KLIH+ELL
Sbjct: 541 KLIHRELL 548
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1504

A DNA sequence (GBSx1591) was identified in *S.agalactiae* <SEQ ID 4625> which encodes the amino acid sequence <SEQ ID 4626>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2692(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC77149 GB:AE000491 orf, hypothetical protein [Escherichia coli K12]
 Identities = 211/363 (58%), Positives = 270/363 (74%), Gaps = 9/363 (2%)

Query: 1   MPNVKDITRESWILSTFPEWGTWLNEEIEEEVVAEGNFAMWWLGNCGVWIKTPGGANVVM 60
           M  VK ITRESWILSTFPEWG+WLNEEIE+E VA G FAMWWLG  G+W+K+ GG NV +
Sbjct: 3   MSKVKSITRESWILSTFPEWGSWLNEEIEQEQVAPGTFAMWWLGCTGIWLKSEGGTNVCV 62

Query: 61  DLWSNRGKSTKKVKDMVRGHQMANMAGVRKLQPNLRAQPMVIDPFAINELDYYLVSHFHS 120
           D W   GK +     M +GHQM  MAGV+KLQPNLR  P V+DPFAI ++D  L +H H+
Sbjct: 63  DFWCGTGKQSHGNPLMKQGHQMQRMAGVKKLQPNLRTTPFVLDPFAIRQIDAVLATHDHN 122

Query: 121 DHIDINTAAAIINNPNLDHVKFVGPYECGEIWKKWGVPEERIIVIKPGESFEFKDIKVTA 180
           DHID+N AAA++ N   D V F+GP  C ++W  WGVP+ER IV+KPG+  + KDI++ A
Sbjct: 123 DHIDVNVAAAVMQNC-ADDVPFIGPKTCVDLWIGWGVPKERCIVVKPGDVVKVKDIEIHA 181

Query: 181 VESFDRTCLVTLPVDGAEEHDGELAGLAVTDEEMARKAVNYIFETPGGTIYHGADSHFSN 240
           +++FDRT L+TLP D       + AG  V  + M  +AVNY+F+TPGG++YH  DSH+SN
Sbjct: 182 LDAFDRTALITLPADQ------KAAG--VLPDGMDDRAVNYLFKTPGGSLYHSGDSHYSN 233

Query: 241 YFAKHGKDYKIDVAINNYGDNPVGIQDKMTSIDLLRMAENLRAKVIIPVHYDIWSNFMAS 300
           Y+AKHG +++IDVA+ +YG+NP GI DKMTS D+LRM E L AKV+IP H+DIWSNF A
Sbjct: 234 YYAKHGNEHQIDVALGSYGENPRGITDKMTSADMLRMGEALNAKVVIPFHHDIWSNFQAD 293

Query: 301 TDEILQLWKMRKERLQYDFHPFIWEVGGKYTYPQDKDRIEYHHPRGFDDCFEQESNIQFK 360
             EI  LW+M+K+RL+Y F PFIW+VGGK+T+P DKD  EYH+PRGFDDCF  E ++ FK
Sbjct: 294 PQEIRVLWEMKKDRLKYGFKPFIWQVGGKFTWPLDKDNFEYHYPRGFDDCFTIEPDLPFK 353

Query: 361 ALL 363
           + L
Sbjct: 354 SFL 356
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4627> which encodes the amino acid sequence <SEQ ID 4628>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3298(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 315/363 (86%), Positives = 348/363 (95%)

Query: 1   MPNVKDITRESWILSTFPEWGTWLNEEIEEEVVAEGNFAMWWLGNCGVWIKTPGGANVVM 60
           M  V+DITRESWIL+TFPEWGTWLNEEIE+EVV   NFAMWWLGNCG+WIKTPGGANVVM
Sbjct: 1   MTKVQDITRESWILNTFPEWGTWLNEEIEQEVVPADNFAMWWLGNCGIWIKTPGGANVVM 60

Query: 61  DLWSNRGKSTKKVKDMVRGHQMANMAGVRKLQPNLRAQPMVIDPFAINELDYYLVSHFHS 120
           DLWSNRGK+TK+VKDMVRGHQMANMAG RKLQPNLRAQPMVIDPF INELDYYLVSH+HS
Sbjct: 61  DLWSNRGKATKQVKDMVRGHQMANMAGARKLQPNLRAQPMVIDPFMINELDYYLVSHYHS 120

Query: 121 DHIDINTAAAIINNPNLDHVKFVGPYECGEIWKKWGVPEERIIVIKPGESFEFKDIKVTA 180
           DHIDINTAAAIINNP L+HVKFVGPYECGE+WK WGVP++RI+++KPG+SFEFKDIK+TA
Sbjct: 121 DHIDINTAAAIINNPKLNHVKFVGPYECGEVWKNWGVPKDRIMILKPGDSFEFKDIKITA 180

Query: 181 VESFDRTCLVTLPVDGAEEHDGELAGLAVTDEEMARKAVNYIFETPGGTIYHGADSHFSN 240
           VESFDRTCLVTLP+ GA+  DG+LAGLA+TD++MARKAVNYIFETPGGTIYHGADSHFSN
Sbjct: 181 VESFDRTCLVTLPIQGADAQDGDLAGLAITDDDMARKAVNYIFETPGGTIYHGADSHFSN 240

Query: 241 YFAKHGKDYKIDVAINNYGDNPVGIQDKMTSIDLLRMAENLRAKVIIPVHYDIWSNFMAS 300
           YFAKHG+DY IDV +NNYG+NP+GIQDKMTS+DLLRMAENLRAKV+IPVHYDIWSNFMAS
Sbjct: 241 YFAKHGRDYDIDVVLNNYGENPIGIQDKMTSVDLLRMAENLRAKVVIPVHYDIWSNFMAS 300

Query: 301 TDEILQLWKMRKERLQYDFHPFIWEVGGKYTYPQDKDRIEYHHPRGFDDCFEQESNIQFK 360
           TDEIL+LWKMRKERLQYDFHPFIWEVGGKYTYPQD++RIEYHHPRGFDDCF ++SNIQFK
Sbjct: 301 TDEILELWKMRKERLQYDFHPFIWEVGGKYTYPQDQNRIEYHHPRGFDDCFLEDSNIQFK 360

Query: 361 ALL 363
           ALL
Sbjct: 361 ALL 363
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1505

A DNA sequence (GBSx1592) was identified in *S.agalactiae* <SEQ ID 4629> which encodes the amino acid sequence <SEQ ID 4630>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3988(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10145> which encodes amino acid sequence <SEQ ID 10146> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA18808 GB:D90917 hypothetical protein [Synechocystis sp.]
 Identities = 358/785 (45%), Positives = 494/785 (62%), Gaps = 15/785 (1%)

Query: 22  LEKLDAWWRAANYISAAQMYLKDNPLLRRELVENDLKVHPIGHWGTVPGQNFIYAHLNRA 81
           L ++  +WRAANY++   +YL+DNPLLR  L    +K   +GHWG+ PG +F+Y HLNR
Sbjct: 44  LNQMHGFWRAANYLAVGMIYLRDNPLLREPLQPEQIKHRLLGHWGSSPGISFLYTHLNRI 103

Query: 82  INKYDLDMFYIEGPGHGGQVMVSNSYLDGSYTELNPNIEQTEDGFKQLCKIFSFPCGIAS 141
           I K+D DM Y+ GPGHG    +   YL+GSY+       + EDG K+  K FSFP GI S
Sbjct: 104 IRKFDQDMLYMVGPGHGAPGFLGPCYLEGSYSRFFAECSEDEDGMKRFFKQFSFPGGIGS 163

Query: 142 HAAPETPGSIHEGGELGYALSHATGAILDNPDVIAATVIGDGEGETGPLMAGWLSNTFIN 201
           H  PETPGSIHEGGELGY LSHA GA  DNP++I   + GDGE ETGPL   W SN FIN
Sbjct: 164 HCTPETPGSIHEGGELGYCLSHAYGAAFDNPNLIVVGLAGDGESETGPLATSWHSNKFIN 223

Query: 202 PVNDGAVLPIFYLNGGKIHNPTIFERKTDEELSQFFEGLGWKPIFADVVELSEDHAAAHA 261
           P+ DGAVLP+ +LNG KI+NP++  R + EEL   FEG G+ P F +      D  + H
Sbjct: 224 PIRDGAVLPVLHLNGYKINNPSVLSRISHEELKALFEGYGYTPYFVE----GSDPESMHQ 279

Query: 262 LFAEKLDQAIQEIKTIQSEARQKPAEEAIQAKFPVLVARIPKGWTGPKAWEGTPIEGGFR 321
             A  LD  + EI  IQ EAR      A++ ++P++V R PKGWTGP   +G  +EG +R
Sbjct: 280 AMAATLDHCVSEIHQIQQEARSTGI--AVRPRWPMVVMRTPKGWTGPDYVDGHKVEGFWR 337

Query: 322 AHQVPIPVDAHHMEHVDSLLSWLQSYRPEELFDENGKIVDEIAAISPKGDRRMSMNPITN 381
           +HQVP+     +  H+  L +W++SY+PEELFDE G +     AI+P+GD+R+   P  N
Sbjct: 338 SHQVPMGGMHENPAHLQQLEAWMRSYKPEELFDEQGTLKPGFKAIAPEGDKRLGSTPYAN 397

Query: 382 AGIV-KAMDTADWKKFALDINVPGQIMAQDMIEFGKYAADLVDANPDNFRIFGPDETKSN 440
            G++ + +   D++++ +D++ PG I A +    G +  D++  N  NFR+FGPDE  SN
Sbjct: 398 GGLLRRGLKMPDFRQYGIDVDQPGTIEAPNTAPLGVFLRDVMANNMTNFRLFGPDENSSN 457

Query: 441 RLQEVFTRTSRQWLGRRKPDYDEA--LSPAGRVIDSQLSEHQAEGFLEGYVLTGRHGFFA 498
           +L  V+  + + W+     +  +   LSP GRV++  LSEH  EG+LE Y+LTGRHGFFA
Sbjct: 458 KLHAVYEVSKKFWIAEYLEEDQDGGELSPDGRVME-MLSEHTLEGWLEAYLLTGRHGFFA 516

Query: 499 SYESFLRVVDSMVTQHFKWLRKSKTHTTWRKNYPALNLIAASTVFQQDHNGYTHQDPGIL 558
           +YESF  V+ SMV QH KWL   + H  WR +  +LN++  STV++QDHNG+THQDPG L
Sbjct: 517 TYESFAHVITSMVNQHAKWLDICR-HLNWRADISSLNILMTSTVWRQDHNGFTHQDPGFL 575

Query: 559 THLAEKTPEYIREYLPADTNSLLAVMDKAFKAEDKINLIVTSKHPRPQFYSIAEAEELVA 618
             +  K+P+ +R YLP D NSLL+V D   ++++ IN+IV  K    Q+  +  A
Sbjct: 576 DVILNKSPDVVRIYLPPDVNSLLSVADHCLQSKNYINIIVCDKQAHLQYQDMTSAIRNCT 635

Query: 619 EGYKVIDWASNVSLNQEPDVVFAAAGTEPNLEALAAISILHKAFPELKIRFVNVLDILKL 678
           +G  + +WASN     EPDVV AAAG  P  EALAA ++L + FP L+IRFV+V+D+LKL
Sbjct: 636 KGVDIWEWASN-DAGTEPDVVMAAAGDIPTKEALAATAMLRQFFPNLRIRFVSVIDLLKL 694

Query: 679 RHPSQDARGLSDEEFNKVFTTDKPVIFAFHGYEDMIRDIFFSRHNH-NLHTHGYRENGDI 737
           +  S+   GLSD +F+ +FTTDKP+IF FH Y  +I  + + R NH NLH  GY+E G+I
Sbjct: 695 QPESEHPHGLSDRDFDSLFTTDKPIIFNFHAYPWLIHRLTYRRTNHGNLHVRGYKEKGNI 754

Query: 738 TTPFDMRVMSELDRFHLAQDA--ALASLGNKAQAFSDEMNQMVAYHKDYIREHGDDIPEV 795
            TP D+ + +++DRF LA D    L  L        + +  M     +Y  EHG D+PE+
Sbjct: 755 NTPMDLAIQNQIDRFSLAIDVIDRLPQLRVAGAHIKEMLKDMQIDCTNYAYEHGIDMPEI 814

Query: 796 QNWKW 800
            NW+W
Sbjct: 815 VNWRW 819
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1506

A DNA sequence (GBSx1593) was identified in *S.agalactiae* <SEQ ID 4631> which encodes the amino acid sequence <SEQ ID 4632>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3509(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF37878 GB:AF234619 OpuAA [Lactococcus lactis]
 Identities = 274/402 (68%), Positives = 338/402 (83%)

Query: 5   LEVKNLTKIFGKKQKAALEMVKQGKSKTEILEKTGATVGVYDASFEIKEGEIFVIMGLSG 64
           +++++LTKIFGK+ K AL MV++G+ K EIL+KTGATVGVYD +FEI EGEIFVIMGLSG
Sbjct: 5   IKIEHLTKIFGKRIKTALTMVEKGEPKNEILKKTGATVGVYDTNFEINEGEIFVIMGLSG 64

Query: 65  SGKSTLVRMLNRLIDPSSGNIYLDGKDIAKMNVEDLRNIRRHDINMVFQNFGLFPHRTIL 124
           SGKSTL+R+LNRLI+P+SG I++D +D+A +N EDL  +RR  ++MVFQNFGLFPHRTIL
Sbjct: 65  SGKSTLLRLLNRLIEPTSGKIFIDNQDVATLNKEDLLQVRRKTMSMVFQNFGLFPHRTIL 124

Query: 125 ENTEFGLEMRGVSKEERTTLAEKALDNAGLLPFKDQYPSQLSGGMQQRVGLARALANSPK 184
           ENTE+GLE++ V KEER   AEKALDNA LL FKDQYP QLSGGMQQRVGLARALAN P+
Sbjct: 125 ENTEYGLEVQNVPKEERRKRAEKALDNANLLDFKDQYPKQLSGGMQQRVGLARALANDPE 184

Query: 185 ILLMDEAFSALDPLIRREMQDELLDLQDTNKQTIIFISHDLNEALRIGDRIALMKDGEIM 244
           ILLMDEAFSALDPLIRREMQDELL+LQ   ++TIIF+SHDLNEALRIGDRIA+MKDG+IM
Sbjct: 185 ILLMDEAFSALDPLIRREMQDELLELQAKFQKTIIFVSHDLNEALRIGDRIAIMKDGKIM 244

Query: 245 QIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTVLEIDGPQVALTRMHREE 304
           QIGTGEEILTNPAND+V+ FVEDVDR+KV+TA+NIMI  LTT +++DGP VAL +M  EE
Sbjct: 245 QIGTGEEILTNPANDYVKTFVEDVDRAKVITAENIMIPALTTNIDVDGPSVALKKMKTEE 304

Query: 305 VSMLMATNRRRQLLGSLTADAAIEARKKDLPLSEVIDKDVVTVSKDTVITDIMPLIYDSS 364
           VS LMA +++RQ  G +T++ AI ARK + PL +V+  DV TVSK+ ++ DI+P+IYD+
Sbjct: 305 VSSLMAVDKKRQFRGVVTSEQAIAARKNNQPLKDVMTTDVGTVSKEMLVRDILPIIYDAP 364

Query: 365 APIAVTDDNDRLLGVIIRGRVIEALANVQDETVVESPKETVE 406
            P+AV DDN  L GV+IRG V+EALA++ DE  VE  ++  E
Sbjct: 365 TPLAVVDDNGFLKGVLIRGSVLEALADIPDEDEVEEIEKEEE 406
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4633> which encodes the amino acid sequence <SEQ ID 4634>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3761(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 344/395 (87%), Positives = 374/395 (94%)

Query: 1   MTNILEVKNLTKIFGKKQKAALEMVKQGKSKTEILEKTGATVGVYDASFEIKEGEIFVIM 60
           M  ILEVK+L+KIFGKKQKAALEMVK GK+K+EI +KTGATVGVYDASFE+K+GEIFVIM
Sbjct: 1   METILEVKHLSKIFGKKQKAALEMVKTGKNKSEIFKKTGATVGVYDASFEVKKGEIFVIM 60

Query: 61  GLSGSGKSTLVRMLNRLIDPSSGNIYLDGKDIAKMNVEDLRNIRRHDINMVFQNFGLFPH 120
```

```
             GLSGSGKSTLVRMLNRLI+PS+G+I L+GKDI+ M+ + LR +RRHDINMVFQ+F LFPH
Sbjct: 61    GLSGSGKSTLVRMLNRLIEPSAGSILLEGKDISTMSADQLREVRRHDINMVFQSFALFPH 120

Query: 121 RTILENTEFGLEMRGVSKEERTTLAEKALDNAGLLPFKDQYPSQLSGGMQQRVGLARALA 180
            +TILENTEFGLE+RGV KEER  LAEKALDN+GLL FKDQYP+QLSGGMQQRVGLARALA
Sbjct: 121 KTILENTEFGLELRGVPKEERQRLAEKALDNSGLLDFKDQYPNQLSGGMQQRVGLARALA 180

Query: 181 NSPKILLMDEAFSALDPLIRREMQDELLDLQDTNKQTIIFISHDLNEALRIGDRIALMKD 240
            NSPKILLMDEAFSALDPLIRREMQDELLDLQD+ KQTIIFISHDLNEALRIGDRIALMKD
Sbjct: 181 NSPKILLMDEAFSALDPLIRREMQDELLDLQDSMKQTIIFISHDLNEALRIGDRIALMKD 240

Query: 241 GEIMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTVLEIDGPQVALTRM 300
            G+IMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTT +E+DGPQVAL RM
Sbjct: 241 GQIMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTTVELDGPQVALNRM 300

Query: 301 HREEVSMLMATNRRRQLLGSLTADAAIEARKKDLPLSEVIDKDVVTVSKDTVITDIMPLI 360
            H EEVSMLMATNRRRQL+GSLTADAAIEARKK LPLSEVID+DV TVSKDT+ITDI+PLI
Sbjct: 301 HNEEVSMLMATNRRRQLVGSLTADAAIEARKKGLPLSEVIDRDVRTVSKDTIITDILPLI 360

Query: 361 YDSSAPIAVTDDNDRLLGVIIRGRVIEALANVQDE 395
            YDSSAPIAVTDDN+RLLGVIIRGRVIEALAN+ DE
Sbjct: 361 YDSSAPIAVTDDNNRLLGVIIRGRVIEALANISDE 395
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1507

A DNA sequence (GBSx1594) was identified in *S.agalactiae* <SEQ ID 4635> which encodes the amino acid sequence <SEQ ID 4636>. This protein is predicted to be OpuABC (opuAB). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.67   Transmembrane   48 -  64 (  43 -  72)
   INTEGRAL    Likelihood = -9.24   Transmembrane  101 - 117 (  93 - 122)
   INTEGRAL    Likelihood = -7.54   Transmembrane  296 - 312 ( 290 - 316)
   INTEGRAL    Likelihood = -6.21   Transmembrane  252 - 268 ( 250 - 273)
   INTEGRAL    Likelihood = -5.57   Transmembrane  141 - 157 ( 138 - 170)
   INTEGRAL    Likelihood = -0.53   Transmembrane  220 - 236 ( 220 - 237)

----- Final Results -----
               bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]
 Identities = 345/578 (59%), Positives = 429/578 (73%), Gaps = 8/578 (1%)

Query: 1   MENLLQHKLPVAPFVESTTNWITKTFSGLFDFIQTIGNALMDWMTKTLLFINPLLFIVLI 60
            M +L   ++P+A +V S T+WIT TFS  FD IQ  G  LM+ +T  L  +   L I ++
Sbjct: 1   MIDLAIGQVPIANWVSSATDWITSTFSSGFDVIQKSGTVLMNGITGALTAVPFWLMIAVV 60

Query: 61  TIAVFFLAKKKWQLPTFTFIGLLFIYNQGLWEQLINTFNLVLVASLISIIIGVPLGIWMA 120
            TI    ++ KK   P FTFIGL  I NQGLW  L++T  LVL++SL+SIIIGVPLGIWMA
Sbjct: 61  TILAILVSGKKIAFPLFTFIGLSLIANQGLWSDLMSTITLVLLSSLLSIIIGVPLGIWMA 120

Query: 121 KSDKVKQVVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVVFALPPTVRFTNLAIR 180
            KSD V ++V PILDFMQTMP FVYLIPAVAFFGIG+VPGVFASV+FALPPTVR TNL IR
Sbjct: 121 KSDLVAKIVQPILDFMQTMPGFVYLIPAVAFFGIGVVPGVFASVIFALPPTVRMTNLGIR 180

Query: 181 EIPLELIEASDSFGSTVKQKLFKVELPLAKNTIMAGINQTMMLALSMVVTGSMIGAPGLG 240
            ++  EL+EA+DSFGST +QKLFK+E PLAK TIMAG+NQT+MLALSMVV  SMIGAPGLG
```

```
Sbjct: 181 QVSTELVEAADSFGSTARQKLFKLEFPLAKGTIMAGVNQTIMLALSMVVIASMIGAPGLG 240

Query: 241 REVLSALQHADIGTGFVSGLSLVILAIVLDRVSQFFNSKPGEKQAKTSKVKKW---VGLG 297
           R VL+A+Q ADIG GFVSG+SLVILAI++DR +Q  N  P EKQ   + VKKW   + L
Sbjct: 241 RGVLAAVQSADIGKGFVSGISLVILAIIIDRFTQKLNVSPLEKQGNPT-VKKWKRGIALV 299

Query: 298 ALALFILAALGRIVVNMTSGNEAKGQKVKIAYVQWDSEVASTNVIAEVLKSKGYDVELTP 357
           +L   I+ A       M+ G  A  +KV + Y+ WDSEVAS NV+ + +K  G+DV+ T
Sbjct: 300 SLLALIIGAFS----GMSFGKTASDKKVDLVYMNWDSEVASINVLTQAMKEHGFDVKTTA 355

Query: 358 LDNAVMWQTVANGNADFTTSAWLPKTHGQYFNKYKNSLDDLGPHVENVKIGLVVPKYMNV 417
           LDNAV WQTVANG AD   SAWLP TH    + KY  S+D LGP+++  K+G VVP YMNV
Sbjct: 356 LDNAVAWQTVANGQADGMVSAWLPNTHKTQWQKYGKSVDLLGPNLKGAKVGFVVPSYMNV 415

Query: 418 NSIEELSNQADKQITGIEPGAGIMKSAKQSLKDYPNLSSWKLLSASTGAMTTTLGKAIKN 477
           NSIE+L+NQA+K ITGIEPGAG+M +++++L  Y NL  WKL+ +S+GAMT  LG+AIK
Sbjct: 416 NSIEDLTNQANKTITGIEPGAGVMAASEKTLNSYDNLKDWKLVPSSSGAMTVALGEAIKQ 475

Query: 478 KDQVVITGWSPHWMFAKYDLKYLKDPKKSFGGEEHINTIARKNLKKDMPKVYKIIDKFKW 537
              +VITGWSPHWMF KYDLKYL DPK + G  E+INTI RK LKK+ P+ YK++DKF W
Sbjct: 476 HKDIVITGWSPHWMFNKYDLKYLADPKGTMGTSENINTIVRKGLKKENPEAYKVLDKFNW 535

Query: 538 TKEDMESIMLDMDKGMEPAKAAQKWIKNHKKEVSEWTK 575
           T +DME++MLD+  G  P +AA+ WIK+H+KEV +W K
Sbjct: 536 TTKDMEAVMLDIQNGKTPEEAAKNWIKDHQKEVDKWFK 573
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4637> which encodes the amino acid sequence <SEQ ID 4638>. Analysis of this protein sequence reveals the following:

```
     Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.86   Transmembrane  101 - 117 (  93 - 121)
   INTEGRAL    Likelihood = -7.54   Transmembrane  252 - 268 ( 250 - 273)
   INTEGRAL    Likelihood = -6.85   Transmembrane   48 -  64 (  43 -  70)
   INTEGRAL    Likelihood = -5.57   Transmembrane  141 - 157 ( 138 - 170)
   INTEGRAL    Likelihood = -5.26   Transmembrane  295 - 311 ( 289 - 315)
   INTEGRAL    Likelihood = -0.53   Transmembrane  220 - 236 ( 220 - 237)

----- Final Results -----
               bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]
 Identities = 340/571 (59%), Positives = 418/571 (72%), Gaps = 8/571 (1%)

Query: 8   KLPVAQLVEQLTEWLTKTFSGLFDIMQVVGSFLMDWMTKTLLFIHPLLFIVLVTAGMFFL 67
           ++P+A  V   T+W+T TFS  FD++Q  G+ LM+ +T  L  +   L I +VT    +
Sbjct: 8   QVPIANWVSSATDWITSTFSSGFDVIQKSGTVLMNGITGALTAVPFWLMIAVVTILAILV 67

Query: 68  AKKKWPLPTFTLLGLLFIYNQGLWKQLMNTFTLVLVASLISVLIGIPLGIWMAKNATVRQ 127
           + KK   P FT +GL  I NQGLW  LM+T TLVL++SL+S++IG+PLGIWMAK+  V +
Sbjct: 68  SGKKIAFPLFTFIGLSLIANQGLWSDLMSTITLVLLSSLLSIIIGVPLGIWMAKSDLVAK 127

Query: 128 IVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVIFALPPTVRFTNLAIRDIPTELI 187
           IV PILDFMQTMP FVYLIPAVAFFGIG+VPGVFASVIFALPPTVR TNL IR + TEL+
Sbjct: 128 IVQPILDFMQTMPGFVYLIPAVAFFGIGVVPGVFASVIFALPPTVRMTNLGIRQVSTELV 187

Query: 188 EASDAFGSTGKQKLFKVELPLAKNTIMAGVNQTMMLALSMVVTGSMIGAPGLGREVLSAL 247
           EA+D+FGST +QKLFK+E PLAK TIMAGVNQT+MLALSMVV  SMIGAPGLGR VL+A+
Sbjct: 188 EAADSFGSTARQKLFKLEFPLAKGTIMAGVNQTIMLALSMVVIASMIGAPGLGRGVLAAV 247

Query: 248 QHADIGSGFVSGLALVILAIVLDRMTQLFNSKPQEKAKAGKTNKW---IGLAALAVFLIA 304
           Q ADIG GFVSG++LVILAI++DR TQ  N  P EK       KW   I L +L   +I
Sbjct: 248 QSADIGKGFVSGISLVILAIIIDRFTQKLNVSPLEKQGNPTVKKWKRGIALVSLLALIIG 307
```

```
Query: 305 ALGRGIMAMTSGMADKGETVNIAYVQWDSEVASTHVIAEVLKNEGYHVTLTPLDNAVMWQ 364
           A       M+ G     + V++ Y+ WDSEVAS +V+ + +K  G+ V  T LDNAV WQ
Sbjct: 308 AFS----GMSFGKTASDKKVDLVYMNWDSEVASINVLTQAMKEHGFDVKTTALDNAVAWQ 363

Query: 365 TVANGNADFSTSAWLPVTHGQQYQKYKSKLDDLGPNLKGTKLGLAVPKYMTDVNSIEDLS 424
           TVANG AD   SAWLP TH  Q+QKY   +D LGPNLKG K+G  VP YM +VNSIEDL+
Sbjct: 364 TVANGQADGMVSAWLPNTHKTQWQKYGKSVDLLGPNLKGAKVGFVVPSYM-NVNSIEDLT 422

Query: 425 KQADQKITGIEPGAGIMAAAQKTLKEYHNLSSWELVAASTGAMTTSLDQAIKKKDPIVVT 484
            QA++ ITGIEPGAG+MAA++KTL  Y NL  W+LV +S+GAMT +L +AIK+   IV+T
Sbjct: 423 NQANKTITGIEPGAGVMAASEKTLNSYDNLKDWKLVPSSSGAMTVALGEAIKQHKDIVIT 482

Query: 485 AWSPHWMFAKYDLKYLKDPKEIFGSTENINTIARKGLKKELPNVYKIIDKFHWTQKDMEA 544
            WSPHWMF KYDLKYL DPK   G++ENINTI RKGLKKE P  YK++DKF+WT KDMEA
Sbjct: 483 GWSPHWMFNKYDLKYLADPKGTMGTSENINTIVRKGLKKENPEAYKVLDKFNWTTKDMEA 542

Query: 545 VMLDINKGMSPEAAAKKWVEANKSKVSSWTK 575
           VMLDI  G +PE AAK W++ ++ +V  W K
Sbjct: 543 VMLDIQNGKTPEEAAKNWIKDHQKEVDKWFK 573
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 439/576 (76%), Positives = 513/576 (88%), Gaps = 2/576 (0%)

Query: 1   MENLLQHKLPVAPFVESTTNWITKTFSGLFDFIQTIGNALMDWMTKTLLFINPLLFIVLI 60
           +E +LQ KLPVA  VE  T W+TKTFSGLFD +Q +G+ LMDWMTKTLLFI+PLLFIVL+
Sbjct: 1   LETILQTKLPVAQLVEQLTEWLTKTFSGLFDIMQVVGSFLMDWMTKTLLFIHPLLFIVLV 60

Query: 61  TIAVFFLAKKKWQLPTFTFIGLLFIYNQGLWEQLINTFNLVLVASLISIIIGVPLGIWMA 120
           T  +FFLAKKKW LPTFT +GLLFIYNQGLW+QL+NTF LVLVASLIS++IG+PLGIWMA
Sbjct: 61  TAGMFFLAKKKWPLPTFTLLGLLFIYNQGLWKQLMNTFTLVLVASLISVLIGIPLGIWMA 120

Query: 121 KSDKVKQVVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVVFALPPTVRFTNLAIR 180
           K+  V+Q+VNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASV+FALPPTVRFTNLAIR
Sbjct: 121 KNATVRQIVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVIFALPPTVRFTNLAIR 180

Query: 181 EIPLELIEASDSFGSTVKQKLFKVELPLAKNTIMAGINQTMMLALSMVVTGSMIGAPGLG 240
           +IP ELIEASD+FGST KQKLFKVELPLAKNTIMAG+NQTMMLALSMVVTGSMIGAPGLG
Sbjct: 181 DIPTELIEASDAFGSTGKQKLFKVELPLAKNTIMAGVNQTMMLALSMVVTGSMIGAPGLG 240

Query: 241 REVLSALQHADIGTGFVSGLSLVILAIVLDRVSQFFNSKPGEKQAKTSKVKKWVGLGALA 300
           REVLSALQHADIG+GFVSGL+LVILAIVLDR++Q FNSKP EK AK  K  KW+GL ALA
Sbjct: 241 REVLSALQHADIGSGFVSGLALVILAIVLDRMTQLFNSKPQEK-AKAGKTNKWIGLAALA 299

Query: 301 LFILAALGRIVVNMTSGNEAKGQKVKIAYVQWDSEVASTNVIAEVLKSKGYDVELTPLDN 360
           +F++AALGR ++ MTSG   KG+ V IAYVQWDSEVAST+VIAEVLK++GY V LTPLDN
Sbjct: 300 VFLIAALGRGIMAMTSGMADKGETVNIAYVQWDSEVASTHVIAEVLKNEGYHVTLTPLDN 359

Query: 361 AVMWQTVANGNADFTTSAWLPKTHGQYFNKYKNSLDDLGPHVENVKIGLVVPKYM-NVNS 419
           AVMWQTVANGNADF+TSAWLP THGQ + KYK+ LDDLGP+++  K+GL VPKYM +VNS
Sbjct: 360 AVMWQTVANGNADFSTSAWLPVTHGQQYQKYKSKLDDLGPNLKGTKLGLAVPKYMTDVNS 419

Query: 420 IEELSNQADKQITGIEPGAGIMKSAKQSLKDYPNLSSWKLLSASTGAMTTTLGKAIKNKD 479
           IE+LS QAD++ITGIEPGAGIM +A+++LK+Y NLSSW+L++ASTGAMTT+L +AIK KD
Sbjct: 420 IEDLSKQADQKITGIEPGAGIMAAAQKTLKEYHNLSSWELVAASTGAMTTSLDQAIKKKD 479

Query: 480 QVVITGWSPHWMFAKYDLKYLKDPKKSFGGEEHINTIARKNLKKDMPKVYKIIDKFKWTK 539
            +V+T WSPHWMFAKYDLKYLKDPK+ FG  E+INTIARK LKK++P VYKIIDKF WT+
Sbjct: 480 PIVVTAWSPHWMFAKYDLKYLKDPKEIFGSTENINTIARKGLKKELPNVYKIIDKFHWTQ 539

Query: 540 EDMESIMLDMDKGMEPAKAAQKWIKNHKKEVSEWTK 575
           +DME++MLD++KGM P  AA+KW++ +K +VS WTK
Sbjct: 540 KDMEAVMLDINKGMSPEAAAKKWVEANKSKVSSWTK 575
```

A related GBS gene <SEQ ID 8827> and protein <SEQ ID 8828> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     -6.57
GvH: Signal Score (-7.5): -5.37
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 6 value: -10.67 threshold:  0.0
   INTEGRAL    Likelihood =-10.67   Transmembrane   48 -  64 (  43 -  72)
   INTEGRAL    Likelihood = -9.24   Transmembrane  101 - 117 (  93 - 122)
   INTEGRAL    Likelihood = -7.54   Transmembrane  296 - 312 ( 290 - 316)
   INTEGRAL    Likelihood = -6.21   Transmembrane  252 - 268 ( 250 - 273)
   INTEGRAL    Likelihood = -5.57   Transmembrane  141 - 157 ( 138 - 170)
   INTEGRAL    Likelihood = -0.53   Transmembrane  220 - 236 ( 220 - 237)
   PERIPHERAL  Likelihood =  2.44     159
 modified ALOM score:   2.63

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00938(322 - 2025 of 2325)
GP|7188801|gb|AAF37879.1|AF234619_2|AF234619(8 - 573 of 573) OpuABC {Lactococcus lactis}
%Match = 44.7
%Identity = 60.2  %Similarity = 75.7
Matches = 342  Mismatches = 136  Conservative Sub.s = 88

255       285       315       345       375       405       435       465
ANVQDETVVESPKETVEA**RGQIILENLLQHKLPVAPFVESTTNWITKTFSGLFDFIQTIGNALMDWMTKTLLFINPLL
                                ::|:| :| | |:||| |||  || ||  |  ||: :|  |  :   |
                         MIDLAIGQVPIANWVSSATDWITSTFSSGFDVIQKSGTVLMNGITGALTAVPFWL
                                  10        20        30        40        50

495       525       555       585       615       645       675       705
FIVLITIAVFFLAKKKWQLPTFTFIGLLFIYNQGLWEQLINTFNLVLVASLISIIIGVPLGIWMAKSDKVKQVVNPILDF
 | ::||   ::: ||  :| |||||| :| |||||  |::|  |||::||:|||||||||||||||| | ::| |||||
MIAVVTILAILVSGKKIAFPLFTFIGLSLIANQGLWSDLMSTITLVLLSSLLSIIIGVPLGIWMAKSDLVAKIVQPILDF
              70        80        90        100       110       120       130

735       765       795       825       855       885       915       945
MQTMPAFVYLIPAVAFFGIGMVPGVFASVVFALPPTVRFTNLAIREIPLELIEASDSFGSTVKQKLFKVELPLAKNTIMA
||||| ||||||||||||||:||||||||:|||||||| ||| ||::  ||:||:|||||| :|||||:|:|||| ||||
MQTMPGFVYLIPAVAFFGIGVVPGVFASVIFALPPTVRMTNLGIRQVSTELVEAADSFGSTARQKLFKLEFPLAKGTIMA
              150       160       170       180       190       200       210

975       1005      1035      1065      1095      1125      1155      1185
GINQTMMLALSMVVTGSMIGAPGLGREVLSALQHADIGTGFVSGLSLVILAIVLDRVSQFFNSKPGEKQAKTSKVKKWVG
|:|||:||||||||  |||||||||| ||:|:| |||| |||||:|||||||::|| :| :|  | |||     ||||
GVNQTIMLALSMVVIASMIGAPGLGRGVLAAVQSADIGKGFVSGISLVILAIIIDRFTQKLNVSPLEKQG-NPTVKKW-K
              230       240       250       260       270       280        290

1215      1245      1275      1305      1335      1365      1395      1425
LGALALFILAALGRIVVNMTSGNEAKGQKVKIAYVQWDSEVASTNVIAEVLKSKGYDVELTPLDNAVMWQTVANGNADFT
 |   : :|| :      |: |  |  :|| : |: ||||||| ||: : :|  |:||: | ||||| ||||||| ||
RGIALVSLLALIIGAFSGMSFGKTASDKKVDLVYMNWDSEVASINVLTQAMKEHGFDVKTTALDNAVAWQTVANGQADGM
                310       320       330       340       350       360       370

1455      1485      1515      1545      1575      1605      1635      1665
TSAWLPKTHGQYFNKYKNSLDDLGPHVENVKIGLVVPKYMNVNSIEELSNQADKQITGIEPGAGIMKSAKQSLKDYPNLS
 ||||| ||    :||  |:| |||:::  |:|:||| ||||||||:|:|||:| |||||||||:| :::::|  | ||
VSAWLPNTHKTQWQKYGKSVDLLGPNLKGAKVGFVVPSYMNVNSIEDLTNQANKTITGIEPGAGVMAASEKTLNSYDNLK
                390       400       410       420       430       440       450

1695      1725      1755      1785      1815      1845      1875      1905
SWKLLSASTGAMTTTLGKAIKNKDQVVITGWSPHWMFAKYDLKYLKDPKKSFGGEEHINTIARKNLKKDMPKVYKIIDKF
```

```
 |||: :|:||||  ||:|||    :|||||||||||  ||||||| ||| : |  |:|||| || |||: |: ||::|||
DWKLVPSSSGAMTVALGEAIKQHKDIVITGWSPHWMFNKYDLKYLADPKGTMGTSENINTIVRKGLKKENPEAYKVLDKF
              470       480       490       500       510       520       530

1935      1965      1995      2025      2055      2085      2115      2145
KWTKEDMESIMLDMDKGMEPAKAAQKWIKNHKKEVSEWTK*YRKKHVSFRACFLM*LKSF*LFNISFILF*YIKSERMKE
 || :|||::|||:  |  | :||: |||:|:||| :| |
NWTTKDMEAVMLDIQNGKTPEEAAKNWIKDHQKEVDKWFK
              550       560       570
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1508

A DNA sequence (GBSx1596) was identified in *S.agalactiae* <SEQ ID 4639> which encodes the amino acid sequence <SEQ ID 4640>. This protein is predicted to be a transposase. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.65   Transmembrane  223 - 239 ( 223 - 240)

----- Final Results -----
               bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10057> which encodes amino acid sequence <SEQ ID 10058> was also identified. A related GBS nucleic acid sequence <SEQ ID 10031> which encodes amino acid sequence <SEQ ID 10032> was also identified. A related GBS nucleic acid sequence <SEQ ID 10801> which encodes amino acid sequence <SEQ ID 10802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA50689 GB:X71844 putative transposase [Clostridium perfringens]
 Identities = 94/364 (25%), Positives = 160/364 (43%), Gaps = 35/364 (9%)

Query: 8   KHKHLTLLDRNDIQSGLDRGETFKAIGLNLLKHPTTIAKEVKRN--KQLRESTKDCLDCP 65
           K+KHL + +R  ++  L  G +       L +   T+  E++R   KQ+++  +  +
Sbjct: 12  KNKHLNMKERMIVEIRLKDGFSAYKNTKELNRPINTVLNEIRRGTTKQIKQGKEFHVYFA 71

Query: 66  LLRKAPYVCNGCPKRRINCGYKKTFYLAKQAQRNYEKLLVESREGIPLNKETFWKIDRVL 125
              +A Y  N   + + N  YK           ++ K +V+        K   W +D  +
Sbjct: 72  DTGEAVYKKN---RLKSNRKYKLL------ECSDFIKYVVDKV------KNDHWSLDACV 116

Query: 126 SNGVKKGQRIYHILKTNDLEVSSSTVYRHIKKGYLSITPIDLPRAVKFKKRRKSTLPPIP 185
                 G+ ++    +    +S+ T+Y ++  G L I  IDLP   K  + +KST
Sbjct: 117 ------GEALHSSRFSPSQIISTKTLYNYVDLGLLPIKNIDLP--AKLHRNKKSTRVRNN 168

Query: 186 KAIKEGRRYEDFIEHM-NQSELNSWLEMDTVIGRIGGK--VLLTFNVAFCNFIFAKLMDS 242
           K  K G    D    + N+ E   W E+D V+G    K  VLLT       +     M S
Sbjct: 169 KK-KLGTSISDRPNSIENREEFGHW-EIDCVLGEKSNKDKVLLTLVERKTRYAIISEMSS 226

Query: 243 KTAIETAKHIQVIKRTLYDNKRDFFELFPVILTDNGGEFARVDDIEIDVCGQSQLFFCDP 302
            + I   K +  IK  L      F E+F  I  DNG EFA + + E+    +++++F  P
Sbjct: 227 HSTISVTKALDKIKEFLGSK---FSEVFKSITADNGSEFADLSEFELKT--KTKVYFTHP 281

Query: 303 NRSDQKARIEKNHTLVRDILPKGTSFDNLTQEDINLALSHINSVKRQALNGKTAYELFSF 362
             S +K   E+++ L+R  +PKG    + + E I+   + +N++ R+ L+ KT  ELF
Sbjct: 282 YSSFEKGTNERHNGLIRRFIPKGKRISDYSLETISFIENWMNTLPRKLLDYKTPEELFEI 341

Query: 363 TYGK 366
```

```
                K
Sbjct: 342 HLDK 345
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1509

A DNA sequence (GBSx1597) was identified in *S.agalactiae* <SEQ ID 4641> which encodes the amino acid sequence <SEQ ID 4642>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.30    Transmembrane   56 -  72 (  48 -  79)
   INTEGRAL    Likelihood = -6.85    Transmembrane   11 -  27 (   6 -  30)
   INTEGRAL    Likelihood = -6.69    Transmembrane  129 - 145 ( 126 - 158)
   INTEGRAL    Likelihood = -6.53    Transmembrane   94 - 110 (  90 - 117)
   INTEGRAL    Likelihood = -1.54    Transmembrane  216 - 232 ( 215 - 232)
   INTEGRAL    Likelihood = -1.22    Transmembrane  147 - 163 ( 147 - 165)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9431> which encodes amino acid sequence <SEQ ID 9432> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07666 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 112/224 (50%), Positives = 150/224 (66%), Gaps = 2/224 (0%)

Query: 8   IKDILWFIIPSLFGVLLLMTPFKYNGMTTVAVSVISKTINQWINAVFPIHYIILLIIFIS 67
           +KD LWF+IPS+ GV L M P + +   T+ V+ ++K +   ++   P    I+L I +
Sbjct: 19  LKDYLWFLIPSIIGVGLFMVPIQKDNAITIPVAFLAKQLQGALDDHLPAILTIMLAIVV- 77

Query: 68  CVLALCYRLFRPSFIEKNDLLKEISDITIFWLIIRLIGLALGLMTVLHIGPEMVWGKETG 127
            VL+    LF+P+   KN LLK +  I   WL++R++G     MT+L +GPE VW + TG
Sbjct: 78  -VLSCVATLFKPNLFMKNGLLKSLFVIHPMWLVVRVLGFIFAFMTLLQLGPEAVWSEGTG 136

Query: 128 GLILFDLIGGLFTIFLAAGFILPFLTEFGLLEFVGVFLTPIMRPFFQLPGRSAVNCVASF 187
            L+L+DL+  LFTIFL AG  LPFL  FGLLE  GV L   MRP F LPGRS+++C+AS+
Sbjct: 137 ALLLYDLLPLLFTIFLFAGLFLPFLLNFGLLELFGVLLNKFMRPVFTLPGRSSIDCLASW 196

Query: 188 VGDGTIGIALTDKQYVEGYYTSREAATISTTFSAVSITFCLXXL 231
           +GDGTIG+ LT+KQY EG+YT REAA ISTTFS VSITF +  L
Sbjct: 197 MGDGTIGVLLTNKQYEEGFYTQREAAVISTTFSVVSITFSIVVL 240
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1510

A DNA sequence (GBSx1599) was identified in *S.agalactiae* <SEQ ID 4643> which encodes the amino acid sequence <SEQ ID 4644>. This protein is predicted to be Na/H antiporter homolog (kefB). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.14   Transmembrane  176 - 192 ( 171 - 203)
   INTEGRAL    Likelihood = -9.34   Transmembrane  353 - 369 ( 348 - 373)
   INTEGRAL    Likelihood = -9.24   Transmembrane    3 -  19 (   1 -  26)
   INTEGRAL    Likelihood = -7.17   Transmembrane  145 - 161 ( 142 - 168)
   INTEGRAL    Likelihood = -7.01   Transmembrane   86 - 102 (  81 - 108)
   INTEGRAL    Likelihood = -6.53   Transmembrane   52 -  68 (  51 -  72)
   INTEGRAL    Likelihood = -5.79   Transmembrane   24 -  40 (  23 -  49)
   INTEGRAL    Likelihood = -5.52   Transmembrane  214 - 230 ( 209 - 233)
   INTEGRAL    Likelihood = -4.04   Transmembrane  260 - 276 ( 258 - 278)
   INTEGRAL    Likelihood = -3.66   Transmembrane  287 - 303 ( 287 - 308)
   INTEGRAL    Likelihood = -2.71   Transmembrane  113 - 129 ( 112 - 129)
   INTEGRAL    Likelihood = -2.66   Transmembrane  332 - 348 ( 330 - 349)

----- Final Results -----
               bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA51756 GB:X73329 Na/H antiporter homolog [Lactococcus lactis]
 Identities = 208/376 (55%), Positives = 285/376 (75%), Gaps = 3/376 (0%)

Query: 1   MHIIIQITIILLASVLATLISKRIGIPAVVGQLLVGIIIGPAMLGLVHQNQVLHVLSEIG 60
           M+ I+Q+TI+L+AS++ATL S+R+ IPAV+GQ+LVGI+I P++LGLVH   VL V+SEIG
Sbjct: 1   MNDILQLTIVLIASLIATLASRRLKIPAVIGQMLVGILIAPSVLGLVHSGHVLEVMSEIG 60

Query: 61  VILLMFLAGLEANFDLLKKYLKPSLLVAITGVIVPMALFYFLTRLFGFQINTAIFYGLVF 120
           VILLMFLAGLE++  +LKK  K S+LVAI GVIVP+ +F  +   FG+ ++T+ FYG+VF
Sbjct: 61  VILLMFLAGLESDLTVLKKNFKASMLVAIGGVIVPLIVFGLVAFSFGYGMSTSFFYGIVF 120

Query: 121 AATSISITVEVLQEYNRVKTDTGAIILGAAVADDVLAVLLLSVFIA--TNGSSSNIGLQI 178
           AATS+SITVEVLQEY ++ T  G+IILGAAV DD+LAVL+LS+F +    GS +++  Q
Sbjct: 121 AATSVSITVEVLQEYGKLSTRAGSIILGAAVVDDILAVLILSIFTSFKNGGSGTHLFFQF 180

Query: 179 IIQLLFFVFLFICMKYLVPALFKLIEKVHFFEKYTILAILICFSLSILADKVGMSSIIGS 238
           +++LLFF FLF+  K L+P  +K ++K+    K TI+A++IC  LS+LAD VGMS++IGS
Sbjct: 181 LLELLFFAFLFVVHK-LIPRFWKFVQKLPIANKNTIVALIICLGLSLLADSVGMSAVIGS 239

Query: 239 FFAGLAIGQTSFVDKVEHKISLLSYTFFIPIFFASIALPLKFDGMMSHLHTILIFTALAV 298
           FFAGLAI QT    K+E   S + Y  FIP+FF  IA+ ++FD ++ H   IL+FT LA+
Sbjct: 240 FFAGLAISQTEVSHKIEEYTSAIGYVIFIPVFFVLIAISVQFDSLIHHPWIILLFTLLAI 299

Query: 299 LSKLIPGYFVGRGFNFSKLESLTIGGGMVSRGEMALIIVQVGLAAKIISSTTYSELVIVV 358
           L+K IP YFVG+    S  ES+ IG GM+SRGEMALI+ Q+GL + II+   YSELVIV+
Sbjct: 300 LTKFIPAYFVGKSNKLSTGESMLIGTGMISRGEMALIVAQIGLTSAIITDEVYSELVIVI 359

Query: 359 ILSTIIAPFILKYSFK 374
           IL+T++APF++K   K
Sbjct: 360 ILATVLAPFLIKLVLK 375
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1511

A DNA sequence (GBSx1600) was identified in *S.agalactiae* <SEQ ID 4645> which encodes the amino acid sequence <SEQ ID 4646>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
```

```
----- Final Results -----
                 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14269 GB:Z99116 ypuA [Bacillus subtilis]
 Identities = 86/319 (26%), Positives = 147/319 (45%), Gaps = 34/319 (10%)

Query: 3   IKKLLFAGLAFILFTLASPAYAASDVQKVIDETYVQPDYVLGYSLNQEQRAQTLQLLNYD 62
            +KK+    LA  +  L  P  + +D    + +  V     LG  L++  + + L  +N
Sbjct: 1   MKKIWIGMLAAAVLLLMVPKVSLADA--AVGDVIV----TLGADLSESDKQKVLDEMNVP 54

Query: 63  ESRDTKVKTLNTSSYAKIMNIADDASIQLY----SSVKIKKLGSNDTLAVNIVTPENITK 118
            ++  T V   N   +  +     +A I       SS+ I K GS     +N+ T  NI+
Sbjct: 55  DNATT-VTVTNKEEHEYLGKYISNAQIGSRAISSSSITIAKKGSG----LNVET-HNISG 108

Query: 119 VTEDMYRNAAVTLGIEHATISVAAPIKVTGESALAGIYYSLE-KNGASVSSENKQLAQEE 177
            +T++MY NA +T G++ A + V AP +V+G +AL G+  + E  +  ++S + KQ+A +E
Sbjct: 109 ITDEMYLNALMTAGVKDAKVYVTAPFEVSGTAALTGLIKAYEVSSDEAISEDVKQVANQE 168

Query: 178 LSTLSGINAENKGKEGYDADKLNVALTDIKSAVAKGGSDLSKDDIRKIVEETLKNYHLDN 237
            L T S +  +  G E   A      +  IK   AK G   +K DI K V++   +  L+
Sbjct: 169 LVTTSEL-GDKIGNENAAA-----LIAKIKEEFAKNGVPDNKADIEKQVDDAASD--LNV 220

Query: 238 AVTENQINLIVNFAVNLSQSNVIKNSDFTNTLNNLKDNIVSKAGSKFKNINVNFNANKAV 297
             +T++Q N +V      S  N +KN+D       + D +  KA  K      +      +
Sbjct: 221 TLTDSQKNQLV------SLFNKMKNADI--DWGQVSDQL-DKAKDKITKFIESDEGKNFI 271

Query: 298 ESGKGFLANIWQQIVNFFQ 316
            +    F  +IW  IV+ F+
Sbjct: 272 QKVIDFFVSIWNAIVSIFK 290
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1512

A repeated DNA sequence (GBSx1602) was identified in *S.agalactiae* <SEQ ID 4647> which encodes the amino acid sequence <SEQ ID 4648>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0603(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15719 GB:Z99122 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 76/138 (55%), Positives = 91/138 (65%), Gaps = 12/138 (8%)

Query: 1   MKLKAVHHIAIIVSDYEKSKDFYVNKLGFEIIRENHRPERHDYKLDLRC-GDIELEIFGN 59
            M LK++HHIAII SDYEKSK FYV+KLGF++I+E +R ER  YKLDL   G   +E+F
Sbjct: 1   MLLKSIHHIAIICSDYEKSKAFYVHKLGFQVIQETYREERGSYKLDLSLNGSYVIELF-- 58

Query: 60  RLDDPEYETPPQRIGRPNWPREACGLRHLAFYVPDVEAYKVELENLGIFVEPIRYDDYTG 119
                  +  PP+R  RP    EA GLRHLAF V  ++    EL   GI  EPIR D  TG
Sbjct: 59  -----SFPDPPERQTRP----EAAGLRHLAFTVGSLDKAVQELHEKGIETEPIRTDPLTG 109
```

```
Query: 120 KKMTFFFDPDGLPLELHE 137
           K+ TFFFDPD LPLEL+E
Sbjct: 110 KRFTFFFDPDQLPLELYE 127
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4649> which encodes the amino acid sequence <SEQ ID 4650>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1205(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 99/137 (72%), Positives = 116/137 (84%)

Query: 1   MKLKAVHHIAIIVSDYEKSKDFYVNKLGFEIIRENHRPERHDYKLDLRCGDIELEIFGNR 60
           MKL A+HH+AIIVSDY  SKDFYVNKLGFEIIREN+RP++HDYKLDL CG IELEIFG
Sbjct: 2   MKLNAIHHVAIIVSDYHLSKDFYVNKLGFEIIRENYRPDKHDYKLDLSCGRIELEIFGKV 61

Query: 61  LDDPEYETPPQRIGRPNWPREACGLRHLAFYVPDVEAYKVELENLGIFVEPIRYDDYTGK 120
             DP Y+ PP+R+  P +  EACGLRHLAF V ++E+Y  +L++LGI VEPIR+DDYTG+
Sbjct: 62  TSDPNYQAPPKRVSEPEFKSEACGLRHLAFRVTNIESYVDDLKSLGIPVEPIRHDDYTGE 121

Query: 121 KMTFFFDPDGLPLELHE 137
           KMTFFFDPDGLPLELHE
Sbjct: 122 KMTFFFDPDGLPLELHE 138
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1513

A DNA sequence (GBSx1603) was identified in *S.agalactiae* <SEQ ID 4651> which encodes the amino acid sequence <SEQ ID 4652>. This protein is predicted to be alpha-amylase. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.62    Transmembrane   14 -  30 (   7 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG41778 GB:AF213261 sortase [Streptococcus gordonii]
 Identities = 136/247 (55%), Positives = 174/247 (70%), Gaps = 2/247 (0%)

Query: 2   RNKKKSHGFFNFVRWLLVVLLIIVGLALVFNKPIRNAFIAHQSNHYQISRVSKKTIEKNK 61
           R  KK     N +  +L V+L++V LAL+FN  IRN  +   +N YQ+S+VSKK IEKNK
Sbjct: 6   RRAKKKRSRRNIILNILSVILLLVALALIFNSSIRNMIMVWHTNKYQVSKVSKKEIEKNK 65

Query: 62  KSKTSYDFSSVKSISTESILSAQTKSHNLPVIGGIAIPDVEINLPIFKGLGNTELSYGAG 121
            SK S++F  V+ +STE++L+AQ K+  LPVIGGIAIP++ +NLPIF GL N  L YGAG
Sbjct: 66  ASKGSFNFEKVEPLSTEAVLNAQWKAQQLPVIGGIAIPELSLNLPIFNGLENAGLYYGAG 125

Query: 122 TMKENQIMGGPNNYALASHHVFGLTGSSKMLFSPLEHAKKGMKVYLTDKSKVYTYTITEI 181
           TMKE Q M G  NYALASHHVFG+TG+++MLFSPL+ AK GMK+YLTDK KVYTY+IT +
```

```
Sbjct: 126 TMKETQEM-GKGNYALASHHVFGITGANEMLFSPLDRAKAGMKIYLTDKEKVYTYSITSV 184

Query: 182 SKVTPEHVEVIDD-TPGKSQLTLVTCTDPEATERIIVHAELEKTGEFSTADESILKAFSK 240
              V PE V+V+DD   G +++TLVTC D  AT R IV   LE    +    + IL  F+K
Sbjct: 185 ENVEPERVDVVDDAADGTAEVTLVTCEDAAATSRTIVKGVLESETPYKETPKKILNYFNK 244

Query: 241 KYNQINL 247
             YNQ+ L
Sbjct: 245 SYNQMQL 251
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4653> which encodes the amino acid sequence <SEQ ID 4654>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.12   Transmembrane   18 -  34 (  13 -  38)
   INTEGRAL    Likelihood = -0.32   Transmembrane   94 - 110 (  94 - 110)

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA73122 GB:M77279 alpha-amylase [unidentified cloning vector]
 Identities = 60/122 (49%), Positives = 85/122 (69%)

Query: 7   RRKIKSMSWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKKQIKKNKEAKS 67
           + K +  +W    L+ +L I+GLAL+FN  IR+ ++ +NS  Y V+K+    +KKN   ++
Sbjct: 4   KEKKRGKNWLINSLLVLLFIIGLALIFNNQIRSWVVQQNSRSYAVSKLKPADVKKNMARET 64

Query: 68  TFDFQAVEPVSTESVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTELIYGAGTMKEE 127
           TFDF +VE +STE+V++AQ   + LPVIG IAIP + INLPIFKGL N  L+ GAGTMKE+
Sbjct: 65  TFDFDSVESLSTEAVMKAQFENKNLPVIGAIAIPSVEINLPIFKGLSNVALLTGAGTMKED 124
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 147/245 (60%), Positives = 192/245 (78%)

Query: 2   RNKKKSHGFFNFVRWLLVVLLIIVGLALVFNKPIRNAFIAHQSNHYQISRVSKKTIEKNK 61
           + K++     ++ R LL+ +L+I+GLAL+FNKPIRN  IA  SN YQ+++VSKK I+KNK
Sbjct: 4   KQKRRKIKSMSWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKKQIKKNK 63

Query: 62  KSKTSYDFSSVKSISTESILSAQTKSHNLPVIGGIAIPDVEINLPIFKGLGNTELSYGAG 121
           ++K+++DF +V+ +STES+L AQ  +  LPVIGGIAIP++ INLPIFKGLGNTEL YGAG
Sbjct: 64  EAKSTFDFQAVEPVSTESVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTELIYGAG 123

Query: 122 TMKENQIMGGPNNYALASHHVFGLTGSSKMLFSPLEHAKKGMKVYLTDKSKVYTYTITEI 181
           TMKE Q+MGG NNY+LASHH+FG+TGSS+MLFSPLE A+ GM +YLTDK K+Y Y I ++
Sbjct: 124 TMKEEQVMGGENNYSLASHHIFGITGSSQMLFSPLERAQNGMSIYLTDKEKIYEYIIKDV 183

Query: 182 SKVTPEHVEVIDDTPGKSQLTLVTCTDPEATERIIVHAELEKTGEFSTADESILKAFSKK 241
             V PE V+VIDDT G  ++TLVTCTD EATERIIV  EL+   +F  A   +LKAF+
Sbjct: 184 FTVAPERVDVIDDTAGLKEVTLVTCTDIEATERIIVKGELKTEYDFDKAPADVLKAFNHS 243

Query: 242 YNQIN 246
           YNQ++
Sbjct: 244 YNQVS 248
```

SEQ ID 4652 (GBS266) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 11; MW 26kDa).

GBS266-His was purified as shown in Figure 205, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1514

A DNA sequence (GBSx1604) was identified in *S.agalactiae* <SEQ ID 4655> which encodes the amino acid sequence <SEQ ID 4656>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4657> which encodes the amino acid sequence <SEQ ID 4658>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 711/819 (86%), Positives = 767/819 (92%)

Query: 1   MQDKNLVDVNLTSEMKTSFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP 60
           MQD+NL+DVNLTSEMKTSFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP
Sbjct: 1   MQDRNLIDVNLTSEMKTSFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP 60

Query: 61  HKKSARITGDVMGKYHPHGDSSIYEAMVRMAQWWSYRHMLVDGHGNFGSMDGDGAAAQRY 120
           HKKSARITGDVMGKYHPHGDSSIYEAMVRMAQWWSYRHMLVDGHGNFGSMDGDGAAAQRY
Sbjct: 61  HKKSARITGDVMGKYHPHGDSSIYEAMVRMAQWWSYRHMLVDGHGNFGSMDGDGAAAQRY 120

Query: 121 TEARMSKIALEMLRDINKNTVDFQDNYDGSEREPLVLPARFPNLLVNGATGIAVGMATNI 180
           TEARMSKIALE+LRDINKNTV+FQDNYDGSEREP+VLPARFPNLLVNGATGIAVGMATNI
Sbjct: 121 TEARMSKIALELLRDINKNTVNFQDNYDGSEREPVVLPARFPNLLVNGATGIAVGMATNI 180

Query: 181 PPHNLGESIDAVKLVMDNPDVTTRELMEVIPGPDFPTGALVMGRSGIHRAYETGKGSIVL 240
           PPHNL ESIDAVK+VM++PD TTRELMEVIPGPDFPTGALVMGRSGIHRAY+TGKGSIVL
Sbjct: 181 PPHNLAESIDAVKMVMEHPDCTTRELMEVIPGPDFPTGALVMGRSGIHRAYDTGKGSIVL 240

Query: 241 RSRTEIETTSNGKERIVVTEFPYGVNKTKVHEHIVRLAQEKRIEGITAVRDESSREGVRF 300
           RSRTEIETT  G+ERIVVTEFPYGVNKTKVHEHIVRLAQEKR+EGITAVRDESSREGVRF
Sbjct: 241 RSRTEIETTQTGRERIVVTEFPYGVNKTKVHEHIVRLAQEKRLEGITAVRDESSREGVRF 300

Query: 301 VIEVRRAASANVILNNLFKLTSLQTNFSFNMLAIEKGVPKILSLRQIIDNYIEHQKEVIV 360
           VIE+RR ASA VILNNLFKLTSLQTNFSFNMLAIE GVPKILSLRQIIDNYI HQKEVI+
Sbjct: 301 VIEIRREASATVILNNLFKLTSLQTNFSFNMLAIENGVPKILSLRQIIDNYISHQKEVII 360

Query: 361 RRTQFDKAKAGARAHILEGLLVALDHLDEVITIIRNSETDTIAQAELMSRFELSERQSQA 420
           RRT+FDK KA ARAHILEGLL+ALDHLDEVI IIRNSETD IAQ ELMSRF+LSERQSQA
Sbjct: 361 RRTRFDKDKAEARAHILEGLLIALDHLDEVIAIIRNSETDVIAQTELMSRFDLSERQSQA 420

Query: 421 ILDMRLRRLTGLERDKIQSEYNDLLALIADLADILAKPERVVTIIKEEMDEVKRKYADAR 480
           ILDMRLRRLTGLERDKIQSEY+DLLALIADL+DILAKPER++TIIKEEMDE+KRKYA+ R
Sbjct: 421 ILDMRLRRLTGLERDKIQSEYDDLLALIADLSDILAKPERIITIIKEEMDEIKRKYANPR 480

Query: 481 RTELMIGEVLSLEDEDLIEEEDVLITLSNKGYIKRLAQDEFRAQKRGGRGIQGTGVNNDD 540
           RTELM+GEVLSLEDEDLIEEEDVLITLSNKGYIKRLAQDEFRAQKRGGRG+QGTGVNNDD
```

```
Sbjct: 481 RTELMVGEVLSLEDEDLIEEEDVLITLSNKGYIKRLAQDEFRAQKRGGRGVQGTGVNNDD 540

Query: 541 FVRELVSTSTHDTVLFFTNLGRVYRLKAYEIPEYGRTAKGLPIVNLLKLDEGETIQTIIN 600
           FVREL+STSTHDT+LFFTN GRVYRLKAYEIPEYGRTAKGLPIVNLLKL++GETIQTIIN
Sbjct: 541 FVRELISTSTHDTLLFFTNFGRVYRLKAYEIPEYGRTAKGLPIVNLLKLEDGETIQTIIN 600

Query: 601 ARKEDVANKYFFFTTQQGIVKRTSVSEFSNIRQNGLRAINLKENDELINVLLIDENEDVI 660
           ARKE+ A K FFFTT+QGIVKRT VSEF+NIRQNGLRA+ LKE D+LINVLL    +D+I
Sbjct: 601 ARKEETAGKSFFFTTKQGIVKRTEVSEFNNIRQNGLRALKLKEGDQLINVLLTSGQDDII 660

Query: 661 IGTRTGYSVRFKVNAVRNMGRTATGVRGVNLREGDKVVGASRIVNGQEVLIITEKGYGKR 720
           IGT +GYSVRF   ++RNMGR+ATGVRGV LRE D+VVGASRI + QEVL+ITE G+GKR
Sbjct: 661 IGTHSGYSVRFNEASIRNMGRSATGVRGVKLREDDRVVGASRIQDNQEVLVITENGFGKR 720

Query: 721 TEASEYPTKGRGGKGIKTANITAKNGPLARLVTINGNEDIMVITDTGVIIRTNVANISQT 780
           T A++YPTKGRGGKGIKTANIT KNG LA LVT++G EDIMVIT+ GVIIRTNVANISQT
Sbjct: 721 TSATDYPTKGRGGKGIKTANITPKNGQLAGLVTVDGTEDIMVITNKGVIIRTNVANISQT 780

Query: 781 GRSTMGVKVMRLDQEAKIVTVALVEQEIEDKSNIEDTKE 819
           GR+T+GVK+M+LD +AKIVT  LV+ E    + I   +E
Sbjct: 781 GRATLGVKIMKLDADAKIVTFTLVQPEDSSIAEINTDRE 819
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1515

A DNA sequence (GBSx1605) was identified in *S.agalactiae* <SEQ ID 4659> which encodes the amino acid sequence <SEQ ID 4660>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA04010 GB:AJ000336 L-lactate dehydrogenase [Streptococcus pneumoniae]
 Identities = 290/329 (88%), Positives = 313/329 (94%), Gaps = 1/329 (0%)

Query: 1   MTATKQHKKVILVGDGAVGSSYAFALVNQGIAQELGIIEIPALFDKAVGDAEDLSHALAF 60
           MT+TKQHKKVILVGDGAVGSSYAFALVNQGIAQELGIIEIP L +KAVGDA DLSHALAF
Sbjct: 1   MTSTKQHKKVILVGDGAVGSSYAFALVNQGIAQELGIIEIPQLHEKAVGDALDLSHALAF 60

Query: 61  TSPKKIYAATYADCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFNGI 120
           TSPKKIYAA Y+DCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGF GI
Sbjct: 61  TSPKKIYAAQYSDCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFKGI 120

Query: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALADKIGVDARSVHAYIMGE 180
           FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA+K+ VDARSVHAYIMGE
Sbjct: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAEKLDVDARSVHAYIMGE 180

Query: 181 HGDSEFAVWSHANVAGVQLEQWLQENRDIDEQGLVDLFISVRDAAYSIINKKGATYYGIA 240
           HGDSEFAVWSHAN+AGV LE++L++ +++ E  L++LF  VRDAAY+IINKKGATYYGIA
Sbjct: 181 HGDSEFAVWSHANIAGVNLEEFLKDTQNVQEAELIELFEGVRDAAYTIINKKGATYYGIA 240

Query: 241 VALARITKAILDDENAVLPLSVYQEGQYGDVKDVFIGQPAIVGAHGIVRPVNIPLNDAEL 300
           VALARITKAILDDENAVLPLSV+QEGQYG V++VFIGQPA+VGAHGIVRPVNIPLNDAE
Sbjct: 241 VALARITKAILDDENAVLPLSVFQEGQYG-VENVFIGQPAVVGAHGIVRPVNIPLNDAET 299

Query: 301 QKMQASAEQLKDIIDEAWKNPEFQEASKN 329
           QKMQASA++L+ IIDEAWKNPEFQEASKN
Sbjct: 300 QKMQASAKELQAIIDEAWKNPEFQEASKN 328
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4661> which encodes the amino acid sequence <SEQ ID 4662>. Analysis of this protein sequence reveals the following:

```
        Possible site: 25

   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.17   Transmembrane  106 - 122 ( 106 - 122)

   ----- Final Results -----
                  bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB81558 GB:U60997 L(+)-lactate dehydrogenase [Streptococcus
           bovis]
 Identities = 278/329 (84%), Positives = 297/329 (89%), Gaps = 2/329 (0%)

Query: 1   MTATKQHKKVILVGDGAVGSSYAFALVTQNIAQELGIIDIFK--EKTQGDAEDLSHALAF 58
           MTATKQHKKVILVGDGAVGSSYAFALV Q IAQELGII+I +   K  GDAEDLSHALAF
Sbjct: 1   MTATKQHKKVILVGDGAVGSSYAFALVNQGIAQELGIIEIPQLFNKAVGDAEDLSHALAF 60

Query: 59  TSPKKIYAADYSDCHDADLVVLTAGAPQKPGETRLDLVEKNLRINKEVVTQIVASGFKGI 118
           TSPKKIYAA Y DC DADLVV+TAGAPQKPGETRLDLV KNL INK +VT++V SGFKGI
Sbjct: 61  TSPKKIYAAKYEDCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTEVVKSGFKGI 120

Query: 119 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAAKIGVDARSVHAYIMGE 178
           FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA K+ VDARSVHAYIMGE
Sbjct: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAEKLDVDARSVHAYIMGE 180

Query: 179 HGDSEFAVWSHANVAGVGLYDWLQANRDIDEQGLVDLFISVRDAAYSIINKKGATFYGIA 238
           HGDSEFAVWSHANVAGV L  +L+  ++++E  LV+LF  VRDAAYSIINKKGATFYGIA
Sbjct: 181 HGDSEFAVWSHANVAGVNLESYLKDVQNVEEAELVELFEGVRDAAYSIINKKGATFYGIA 240

Query: 239 VALARITKAILDDENAVLPLSVFQEGQYEGVEDCYIGQPAIVGAYGIVRPVNIPLNDAEL 298
           VALARITKAIL+DENAVLPLSVFQEGQY  V DCYIGQPAIVGA+GIVRPVNIPLNDAE
Sbjct: 241 VALARITKAILNDENAVLPLSVFQEGQYANVTDCYIGQPAIVGAHGIVRPVNIPLNDAEQ 300

Query: 299 QKMQASANQLKAIIDEAFAKEEFASAAKN 327
           QKM+ASA +LKAIIDEAF+KEEFASA KN
Sbjct: 301 QKMEASAKELKAIIDEAFSKEEFASACKN 329
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 286/329 (86%), Positives = 299/329 (89%), Gaps = 2/329 (0%)

Query: 1   MTATKQHKKVILVGDGAVGSSYAFALVNQGIAQELGIIEIPALFDKAVGDAEDLSHALAF 60
           MTATKQHKKVILVGDGAVGSSYAFALV Q IAQELGII+I    +K  GDAEDLSHALAF
Sbjct: 1   MTATKQHKKVILVGDGAVGSSYAFALVTQNIAQELGIIDI--FKEKTQGDAEDLSHALAF 58

Query: 61  TSPKKIYAATYADCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFNGI 120
           TSPKKIYAA Y+DC DADLVV+TAGAPQKPGETRLDLV KNL INK +VTQ+V SGF GI
Sbjct: 59  TSPKKIYAADYSDCHDADLVVLTAGAPQKPGETRLDLVEKNLRINKEVVTQIVASGFKGI 118

Query: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALADKIGVDARSVHAYIMGE 180
           FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA KIGVDARSVHAYIMGE
Sbjct: 119 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAAKIGVDARSVHAYIMGE 178

Query: 181 HGDSEFAVWSHANVAGVQLEQWLQENRDIDEQGLVDLFISVRDAAYSIINKKGATYYGIA 240
           HGDSEFAVWSHANVAGV L  WLQ NRDIDEQGLVDLFISVRDAAYSIINKKGAT+YGIA
Sbjct: 179 HGDSEFAVWSHANVAGVGLYDWLQANRDIDEQGLVDLFISVRDAAYSIINKKGATFYGIA 238

Query: 241 VALARITKAILDDENAVLPLSVYQEGQYGDVKDVFIGQPAIVGAHGIVRPVNIPLNDAEL 300
           VALARITKAILDDENAVLPLSV+QEGQY  V+D +IGQPAIVGA+GIVRPVNIPLNDAEL
Sbjct: 239 VALARITKAILDDENAVLPLSVFQEGQYEGVEDCYIGQPAIVGAYGIVRPVNIPLNDAEL 298
```

```
Query: 301 QKMQASAEQLKDIIDEAWKNPEFQEASKN 329
           QKMQASA QLK IIDEA+   EF  A+KN
Sbjct: 299 QKMQASANQLKAIIDEAFAKEEFASAAKN 327
```

SEQ ID 4660 (GBS312) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 7; MW 40kDa).

GBS312-His was purified as shown in Figure 205, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1516

A DNA sequence (GBSx1606) was identified in *S.agalactiae* <SEQ ID 4663> which encodes the amino acid sequence <SEQ ID 4664>. This protein is predicted to be NADH oxidase (nox). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1888(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC26485 GB:AF014458 NADH oxidase [Streptococcus pneumoniae]
          (ver 2)
 Identities = 363/458 (79%), Positives = 408/458 (88%), Gaps = 3/458 (0%)

Query: 1   MSKIVVVGTNHAGTAAIKTMLSNYGEANEIVTFDQNSNISFLGCGMALWIGEQIDGPEGL 60
           MSKIVVVG NHAGTA I TML N+G  NEIV FDQNSNISFLGCGMALWIGEQIDG EGL
Sbjct: 1   MSKIVVVGANHAGTACINTMLDNFGNENEIVVFDQNSNISFLGCGMALWIGEQIDGAEGL 60

Query: 61  FYSDKEQLESMGAKVYMNSPVLNIDYDKKEVTALVDGKEHVESYEKLILATGSQPIIPPI 120
           FYSDKE+LE+ GAKVYMNSPVL+IDYD K VTA V+GKEH ESYEKLI ATGS PI+PPI
Sbjct: 61  FYSDKEKLEAKGAKVYMNSPVLSIDYDNKVVTAEVEGKEHKESYEKLIFATGSTPILPPI 120

Query: 121 KGVEIQEGSREFKATLENLQFVKLYQNSEEVIEKLAKPG--INRVAVVGAGYIGVELAEA 178
           +GVEI +G+REFKATLEN+QFVKLYQN+EEVI KL+     ++R+AVVG GYIGVELAEA
Sbjct: 121 EGVEIVKGNREFKATLENVQFVKLYQNAEEVINKLSDKSQHLDRIAVVGGGYIGVELAEA 180

Query: 179 FQRIGKEVTLVDVADTCMGGYYDRDFTDMMSKNLEDHGIRLAFGQAVQAVEGDGKVERLV 238
           F+R+GKEV LVD+ DT + GYYD+DFT MM+KNLEDH IRLA GQ V+A+EGDGKVERL+
Sbjct: 181 FERLGKEVVLVDIVDTVLNGYYDKDFTQMMAKNLEDHNIRLALGQTVKAIEGDGKVERLI 240

Query: 239 TDKETFDVDMVILAVGFRPNTELGAGKLDTFRNGAWVVDKKQETSVKDVYAIGDCATIWD 298
           TDKE+FDVDMVILAVGFRPNT L  GK++ FRNGA++VDKKQETS+  VYA+GDCAT++D
Sbjct: 241 TDKESFDVDMVILAVGFRPNTALADGKIELFRNGAFLVDKKQETSIPGVYAVGDCATVYD 300

Query: 299 NSRDDINYIALASNAVRTGIVAAHNACGTELEGAGVQGSNGISIYGLNMVSTGLTLEKAK 358
           N+R D +YIALASNAVRTGIV A+NACG ELEG GVQGSNGISIYGL+MVSTGLTLEKAK
Sbjct: 301 NARKDTSYIALASNAVRTGIVGAYNACGHELEGIGVQGSNGISIYGLHMVSTGLTLEKAK 360

Query: 359 QAGYNAVETGFNDLQKPEFIKHNNHEVAIKIVYDKDSRVILGCQMVSHE-DVSMGIHMFS 417
            AGYNA ETGFNDLQKPEF+KH+NHEVAIKIV+DKDSR ILG QMVSH+  +SMGIHMFS
Sbjct: 361 AAGYNATETGFNDLQKPEFMKHDNHEVAIKIVFDKDSREILGAQMVSHDIAISMGIHMFS 420

Query: 418 LAIQEKVTIEKLALTDIFFLPHFNKPYNYITMAALGAK 455
           LAIQE VTI+KLALTD+FFLPHFNKPYNYITMAAL A+
Sbjct: 421 LAIQEHVTIDKLALTDLFFLPHFNKPYNYITMAALTAE 458
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4665> which encodes the amino acid sequence <SEQ ID 4666>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2068(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 362/456 (79%), Positives = 403/456 (87%)

Query: 1   MSKIVVVGTNHAGTAAIKTMLSNYGEANEIVTFDQNSNISFLGCGMALWIGEQIDGPEGL 60
           MSKIVVVG NHAGTA IKTML+NYG+ANEIV FDQNSNISFLGCGMALWIGEQI GPEGL
Sbjct: 1   MSKIVVVGANHAGTACIKTMLTNYGDANEIVVFDQNSNISFLGCGMALWIGEQIAGPEGL 60

Query: 61  FYSDKEQLESMGAKVYMNSPVLNIDYDKKEVTALVDGKEHVESYEKLILATGSQPIIPPI 120
           FYSDKE+LES+GAKVYM SPV +IDYD K VTALVDGK HVE+Y+KLI ATGSQPI+PPI
Sbjct: 61  FYSDKEELESLGAKVYMESPVQSIDYDAKTVTALVDGKNHVETYDKLIFATGSQPILPPI 120

Query: 121 KGVEIQEGSREFKATLENLQFVKLYQNSEEVIEKLAKPGINRVAVVGAGYIGVELAEAFQ 180
           KG EI+EGS EF+ATLENLQFVKLYQNS +VI KL    I RVAVVGAGYIGVELAEAFQ
Sbjct: 121 KGAEIKEGSLEFEATLENLQFVKLYQNSADVIAKLENKDIKRVAVVGAGYIGVELAEAFQ 180

Query: 181 RIGKEVTLVDVADTCMGGYYDRDFTDMMSKNLEDHGIRLAFGQAVQAVEGDGKVERLVTD 240
           R GKEV L+DV DTC+ GYYDRD TD+M+KN+E+HGI+LAFG+ V+ V G+GKVE+++TD
Sbjct: 181 RKGKEVVLIDVVDTCLAGYYDRDLTDLMAKNMEEHGIQLAFGETVKEVAGNGKVEKIITD 240

Query: 241 KETFDVDMVILAVGFRPNTELGAGKLDTFRNGAWVVDKKQETSVKDVYAIGDCATIWDNS 300
           K  +DVDMVILAVGFRPNT LG GK+D FRNGA++V+K+QETS+  VYAIGDCATI+DN+
Sbjct: 241 KNEYDVDMVILAVGFRPNTTLGNGKIDLFRNGAFLVNKRQETSIPGVYAIGDCATIYDNA 300

Query: 301 RDDINYIALASNAVRTGIVAAHNACGTELEGAGVQGSNGISIYGLNMVSTGLTLEKAKQA 360
             D NYIALASNAVRTGIVAAHNACGT+LEG GVQGSNGISIYGL+MVSTGLTLEKAK+
Sbjct: 301 TRDTNYIALASNAVRTGIVAAHNACGTDLEGIGVQGSNGISIYGLHMVSTGLTLEKAKRL 360

Query: 361 GYNAVETGFNDLQKPEFIKHNNHEVAIKIVYDKDSRVILGCQMVSHEDVSMGIHMFSLAI 420
           G++A  T + D QKPEFI+H N  V IKIVYDKDSR ILG QM + EDVSMGIHMFSLAI
Sbjct: 361 GFDAAVTEYTDNQKPEFIEHGNFPVTIKIVYDKDSRRILGAQMAAREDVSMGIHMFSLAI 420

Query: 421 QEKVTIEKLALTDIFFLPHFNKPYNYITMAALGAKD 456
           QE VTIEKLALTDIFFLPHFNKPYNYITMAALGAKD
Sbjct: 421 QEGVTIEKLALTDIFFLPHFNKPYNYITMAALGAKD 456
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1517

A DNA sequence (GBSx1607) was identified in *S.agalactiae* <SEQ ID 4667> which encodes the amino acid sequence <SEQ ID 4668>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2319(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1518

A DNA sequence (GBSx1608) was identified in *S.agalactiae* <SEQ ID 4669> which encodes the amino acid sequence <SEQ ID 4670>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.75    Transmembrane  160 - 176 ( 157 - 179)
   INTEGRAL    Likelihood = -7.38    Transmembrane   73 -  89 (  70 -  97)
   INTEGRAL    Likelihood = -5.47    Transmembrane  289 - 305 ( 284 - 312)
   INTEGRAL    Likelihood = -4.09    Transmembrane  107 - 123 ( 106 - 124)
   INTEGRAL    Likelihood = -3.24    Transmembrane   43 -  59 (  43 -  59)
   INTEGRAL    Likelihood = -1.91    Transmembrane  258 - 274 ( 258 - 275)
   INTEGRAL    Likelihood = -1.33    Transmembrane  234 - 250 ( 233 - 251)
   INTEGRAL    Likelihood = -0.00    Transmembrane  209 - 225 ( 209 - 225)

----- Final Results -----
               bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9805> which encodes amino acid sequence <SEQ ID 9806> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 172/318 (54%), Positives = 234/318 (73%)

Query: 5   LSLTTIFALLFSSMLIYATPLIFTSIGGTFSERGGIVNVGLEGIMVIGAFSGVVFNLEFA 64
           + +  I +++  + L+YA PLI T++GG FSER G+VN+GLEG+M+IGAF+ V+FNL F
Sbjct: 1   MDIVQILSIIVPATLVYAAPLILTALGGVFSERSGVVNIGLEGLMIIGAFTSVLFNLFFG 60

Query: 65  SVFGDATPWISVLVGGLVGLIFSVIHAVATVNFRADHIISGTVLNLMAPSLAVFLIKVLY 124
              G A PW+S+L     G +FS+IHA A ++FRAD  +SG  +N++A    +F++K++Y
Sbjct: 61  QELGAAAPWLSLLAAMAAGALFSLIHAAAAISFRADQTVSGVAINMLALGATLFIVKLIY 120

Query: 125 NKGQTDNIQESFGKFNFPILSDIPFVGDIFFKGTSLVGYIAILFSFLAWFILYKTRFGLR 184
            K QTD I E F K   P L DIP +G IFF        +AI  +F++WFIL+KT FGLR
Sbjct: 121 GKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSDVYYTSILAIALAFISWFILFKTPFGLR 180

Query: 185 LRSVGEHPQAADTLGINVYLMRYSGVLISGFLGGIGGAVYAQSISVNFAATTILGPGFIS 244
           +RSVGEHP AADT+GINVY MRY GV+ISG  GG+GG VYA +I+++F  +TI G GFI+
Sbjct: 181 IRSVGEHPMAADTMGINVYKMRYIGVMISGLFGGLGGGVYASTIALDFTHSTISGQGFIA 240

Query: 245 LAAMIFGKWNPIGAMLASLFFGLSQSLAVIGSHLPLLSNIPTVYLQIAPYVLTIIVLAAF 304
           LAA++FGKW+PIGA+ A+LFFG +QSL++IGS LPL  +IP VY+ +APY+LTI+ L  F
Sbjct: 241 LAALVFGKWHPIGALGAALFFGFAQSLSIIGSLLPLFKDIPNVYMLMAPYILTILALTGF 300

Query: 305 FGQAVAPKADGINYIKTK 322
            G+A APKA+G+ YIK K
Sbjct: 301 IGRADAPKANGVPYIKGK 318
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4671> which encodes the amino acid sequence <SEQ ID 4672>. Analysis of this protein sequence reveals the following:

```
     Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.92    Transmembrane   73 -  89 (  69 -  97)
```

```
    INTEGRAL    Likelihood = -5.04   Transmembrane  160 - 176 ( 158 - 177)
    INTEGRAL    Likelihood = -4.62   Transmembrane  289 - 305 ( 284 - 312)
    INTEGRAL    Likelihood = -3.98   Transmembrane  234 - 250 ( 232 - 251)
    INTEGRAL    Likelihood = -2.13   Transmembrane  107 - 123 ( 106 - 123)
    INTEGRAL    Likelihood = -2.02   Transmembrane   43 -  59 (  43 -  59)
    INTEGRAL    Likelihood = -0.53   Transmembrane  258 - 274 ( 258 - 274)

----- Final Results -----
               bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 176/318 (55%), Positives = 239/318 (74%)

Query: 5   MSLVTIFALLMSSMLIYATPLIFTSIGGTFSERSGVVNVGLEGIMVMGAFSGIVFNLEFA 64
            M +V I ++++ + L+YA PLI T++GG FSERSGVVN+GLEG+M++GAF+ ++FNL F
Sbjct: 1   MDIVQILSIIVPATLVYAAPLILTALGGVFSERSGVVNIGLEGLMIIGAFTSVLFNLFFG 60

Query: 65  ETFGKATPWIAVLVGGIVGLIFSLIHAVATINFRADHIVSGTVLNLLAPSFAVFLVKAMY 124
            +  G A PW+++L     G +FSLIHA A I+FRAD  VSG  +N+LA    +F+VK +Y
Sbjct: 61  QELGAAAPWLSLLAAMAAGALFSLIHAAAAISFRADQTVSGVAINMLALGATLFIVKLIY 120

Query: 125 GKGQTDNIQQSFGKFDFPGLSQIPVIGDIFFKNTSLIGYFAIAFSFFAWFLLYKTRFGLR 184
            GK QTD I + F K   PGL  IPV+G IFF +        AIA +F +WF+L+KT FGLR
Sbjct: 121 GKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSDVYYTSILAIALAFISWFILFKTPFGLR 180

Query: 185 LRSVGEHPQAADTLGINVYLMKYYGVMISGFLGGIGGAVYAQSISVNFAVTTILGPGFIA 244
            +RSVGEHP AADT+GINVY M+Y GVMISG  GG+GG VYA +I+++F  +TI G GFIA
Sbjct: 181 IRSVGEHPMAADTMGINVYKMRYIGVMISGLFGGLGGGVYASTIALDFTHSTISGQGFIA 240

Query: 245 LAAMIFGKWNPVGAMLSSLFFGLSQSLAVIGAQLPLLEKIPTVYLQIAPYMVTIIILAAF 304
            LAA++FGKW+P+GA+ ++LFFG +QSL++IG+ LPL + IP VY+ +APY++TI+ L  F
Sbjct: 241 LAALVFGKWHPIGALGAALFFGFAQSLSIIGSLLPLFKDIPNVYMLMAPYILTILALTGF 300

Query: 305 FGQAVAPKADGINYIKSK 322
             G+A APKA+G+ YIK K
Sbjct: 301 IGRADAPKANGVPYIKGK 318
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 272/322 (84%), Positives = 301/322 (93%)

Query: 1   MVSKLSLTTIFALLFSSMLIYATPLIFTSIGGTFSERGGIVNVGLEGIMVIGAFSGVVFN 60
            +V+K+SL TIFALL SSMLIYATPLIFTSIGGTFSER G+VNVGLEGIMV+GAFSG+VFN
Sbjct: 1   VVNKMSLVTIFALLMSSMLIYATPLIFTSIGGTFSERSGVVNVGLEGIMVMGAFSGIVFN 60

Query: 61  LEFASVFGDATPWISVLVGGLVGLIFSVIHAVATVNFRADHIISGTVLNLMAPSLAVFLI 120
            LEFA  FG ATPWI+VLVGG+VGLIFS+IHAVAT+NFRADHI+SGTVLNL+APS AVFL+
Sbjct: 61  LEFAETFGKATPWIAVLVGGIVGLIFSLIHAVATINFRADHIVSGTVLNLLAPSFAVFLV 120

Query: 121 KVLYNKGQTDNIQESFGKFNFPILSDIPFVGDIFFKGTSLVGYIAILFSFLAWFILYKTR 180
            K +Y KGQTDNIQ+SFGKF+FP LS IP +GDIFFK TSL+GY AI FSF AWF+LYKTR
Sbjct: 121 KAMYGKGQTDNIQQSFGKFDFPGLSQIPVIGDIFFKNTSLIGYFAIAFSFFAWFLLYKTR 180

Query: 181 FGLRLRSVGEHPQAADTLGINVYLMRYSGVLISGFLGGIGGAVYAQSISVNFAATTILGP 240
            FGLRLRSVGEHPQAADTLGINVYLM+Y GV+ISGFLGGIGGAVYAQSISVNFA TTILGP
Sbjct: 181 FGLRLRSVGEHPQAADTLGINVYLMKYYGVMISGFLGGIGGAVYAQSISVNFAVTTILGP 240

Query: 241 GFISLAAMIFGKWNPIGAMLASLFFGLSQSLAVIGSHLPLLSNIPTVYLQIAPYVLTIIV 300
            GFI+LAAMIFGKWNP+GAML+SLFFGLSQSLAVIG+ LPLL  IPTVYLQIAPY++TII+
Sbjct: 241 GFIALAAMIFGKWNPVGAMLSSLFFGLSQSLAVIGAQLPLLEKIPTVYLQIAPYMVTIII 300

Query: 301 LAAFFGQAVAPKADGINYIKTK 322
            LAAFFGQAVAPKADGINYIK+K
Sbjct: 301 LAAFFGQAVAPKADGINYIKSK 322
```

A related GBS gene <SEQ ID 8829> and protein <SEQ ID 8830> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:      8.61
GvH: Signal Score (-7.5): -1.53
     Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 8 value:  -7.75 threshold:  0.0
   INTEGRAL    Likelihood = -7.75   Transmembrane  160 - 176 ( 157 - 179)
   INTEGRAL    Likelihood = -7.38   Transmembrane   73 -  89 (  70 -  97)
   INTEGRAL    Likelihood = -5.47   Transmembrane  289 - 305 ( 284 - 312)
   INTEGRAL    Likelihood = -4.09   Transmembrane  107 - 123 ( 106 - 124)
   INTEGRAL    Likelihood = -3.24   Transmembrane   43 -  59 (  43 -  59)
   INTEGRAL    Likelihood = -1.91   Transmembrane  258 - 274 ( 258 - 275)
   INTEGRAL    Likelihood = -1.33   Transmembrane  234 - 250 ( 233 - 251)
   INTEGRAL    Likelihood = -0.00   Transmembrane  209 - 225 ( 209 - 225)
   PERIPHERAL  Likelihood =  3.34     139
 modified ALOM score:   2.05

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00914(313 - 1266 of 1566)
EGAD|108729|BS3151(1 - 318 of 319) hypothetical protein {Bacillus subtilis}
GP|1934814|emb|CAB07939.1||Z93937      unknown      {Bacillus       subtilis}
GP|2635653|emb|CAB15146.1||Z99120 similar to hypothetical proteins {Bacillus subtilis}
PIR|F70009|F70009 conserved hypothetical protein yufQ - Bacillus subtilis
%Match = 34.9
%Identity = 54.1  %Similarity = 76.4
Matches = 172  Mismatches = 75  Conservative Sub.s = 71

216       246       276       306       336       366       396       426
TLQVFHLS*LKL*QLQSSSS*VSITLLSMLLNLKNK*KVVSKLSLTTIFALLFSSMLIYATPLIFTSIGGTFSERGGIVN
                                          : :  |::::  : |:|| |||:|::|| |||| |:||
                                          MDIVQILSIIVPATLVYAAPLILTALGGVFSERSGVVN
                                                  10        20        30

456       486       516       546       576       606       636       666
VGLEGIMVIGAFSGVVFNLEFASVFGDATPWISVLVGGLVGLIFSVIHAVATVNFRADHIISGTVLNLMAPSLAVFLIKV
:||||:|:||||: |:||| |   :| | ||:|:|     | :||:||| | ::||||: :||  :|::|    :|::|:
IGLEGLMIIGAFTSVLFNLFFGQELGAAAPWLSLLAAMAAGALFSLIHAAAAISFRADQTVSGVAINMLALGATLFIVKL
          50        60        70        80        90        100       110

696       726       756       786       816       846       876       906
LYNKGQTDNIQESFGKFNFPILSDIPFVGDIFFKGTSLVGYIAILFSFLAWFILYKTRFGLRLRSVGEHPQAADTLGINV
:| | ||| | | | |   | | ||| :| |||        :|| ::|::||||:|| ||||:||||||| ||||:||||
IYGKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSDVYYTSILAIALAFISWFILFKTPFGLRIRSVGEHPMAADTMGINV
          130       140       150       160       170       180       190

936       966       996       1026      1056      1086      1116      1146
YLMRYSGVLISGFLGGIGGAVYAQSISVNFAATTILGPGFISLAAMIFGKWNPIGAMLASLFFGLSQSLAVIGSHLPLLS
| ||| ||:|||::||:|| ||| :|:::|  :|| | |||:|||::||||:||||: |:||||::|||::||| |||:
YKMRYIGVMISGLFGGLGGGVYASTIALDFTHSTISGQGFIALAALVFGKWHPIGALGAALFFGFAQSLSIIGSLLPLFK
          210       220       230       240       250       260       270

1176      1206      1236      1266      1296      1326      1356      1386
NIPTVYLQIAPYVLTIIVLAAFFGQAVAPKADGINYIKTK*IKRN*YKLVSFYCL*ICEKILCENFT*IIIQ*Q*NIKK*
:|| ||: :|||:|||: |  | |:| ||||:|: ||| |
DIPNVYMLMAPYILTILALTGFIGRADAPKANGVPYIKGKR
```

```
          290       300       310
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1519

A DNA sequence (GBSx1609) was identified in *S.agalactiae* <SEQ ID 4673> which encodes the amino acid sequence <SEQ ID 4674>. This protein is predicted to be ribose/galactose ABC transporter, permease protein (rbsC-1). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.59   Transmembrane  205 - 221 ( 200 - 228)
   INTEGRAL    Likelihood =-13.69   Transmembrane   21 -  37 (  13 -  45)
   INTEGRAL    Likelihood = -7.27   Transmembrane  302 - 318 ( 290 - 321)
   INTEGRAL    Likelihood = -7.17   Transmembrane  115 - 131 ( 111 - 138)
   INTEGRAL    Likelihood = -4.25   Transmembrane  251 - 267 ( 250 - 268)
   INTEGRAL    Likelihood = -2.97   Transmembrane   63 -  79 (  63 -  80)
   INTEGRAL    Likelihood = -2.87   Transmembrane  333 - 349 ( 328 - 349)

----- Final Results -----
              bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8831> which encodes amino acid sequence <SEQ ID 8832> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length of  UR:    24
     Peak Value of UR:   3.06
     Net Charge of CR: 3
McG: Discrim Score:     12.53
GvH: Signal Score (-7.5): -5.31
     Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 7 value: -14.59 threshold:  0.0
   INTEGRAL    Likelihood =-14.59   Transmembrane  196 - 212 ( 191 - 219)
   INTEGRAL    Likelihood =-13.69   Transmembrane   12 -  28 (   4 -  36)
   INTEGRAL    Likelihood = -7.27   Transmembrane  293 - 309 ( 281 - 312)
   INTEGRAL    Likelihood = -7.17   Transmembrane  106 - 122 ( 102 - 129)
   INTEGRAL    Likelihood = -4.25   Transmembrane  242 - 258 ( 241 - 259)
   INTEGRAL    Likelihood = -2.97   Transmembrane   54 -  70 (  54 -  71)
   INTEGRAL    Likelihood = -2.87   Transmembrane  324 - 340 ( 319 - 340)
   PERIPHERAL  Likelihood =  0.16     133
 modified ALOM score:   3.42
icml HYPID: 7  CFP: 0.684

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15145 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 154/349 (44%), Positives = 220/349 (62%), Gaps = 6/349 (1%)

Query: 10  MSKKAQKIAVPLISVVLGIILGAIIMLIFGYDPLWGYEGLFQTAFGSIKNIGEIFRAMGP 69
```

```
           M K+   + VPLI+++LG+  GA+IML+ GY    GY  L+   FG I  +GE  R + P
Sbjct: 1   MVKRLSHLLVPLIAIILGLAAGALIMLVSGYSVASGYSALWNGIFGEIYYVGETIRQITP 60

Query: 70  LILIALGFSVASRAGFFNIGLPGQALSGWIAAGWFALSHPDMPRPAMILCTIIIGIVAGG 129
            IL  L  + A R G FNIG+ GQ L GW AA W   +  D P    +   +I    AGG
Sbjct: 61  YILSGLAVAFAFRTGLFNIGVEGQLLVGWTAAVWVGTAF-DGPAYIHLPLALITAAAAGG 119

Query: 130 ITGAIPGILRAYLGTSEVIVTIMMNYIVLYSGNAIVQRVFPKSIMRTSDSSVYVSANASY 189
           + G IPGIL+A     EVIVTIMMNYI L+  N I+  V         D +  +  +AS
Sbjct: 120 LWGFIPGILKARFYVHEVIVTIMMNYIALHMTNYIISNVLTDH----QDKTGKIHESASL 175

Query: 190 QTDWLSSLTNNSRINIGIFIAIIAVVLVWFLLNKTTLGFEIRSVGLNPNASEYAGMSAKR 249
           ++ +L  +T+ SR+++GI +A++A V++WF++NK+T GFE+R+VG N +AS+YAGMS ++
Sbjct: 176 RSPFLEQITDYSRLHLGIIVALLAAVIMWFIINKSTKGFELRAVGFNQHASQYAGMSVRK 235

Query: 250 TIILSMIISGAFAGLGGVVEGLGTFENVFVQPSSLAIGFDGMAVSLLAANSPIGILFAAF 309
            I+ SM+ISGAFAGL G +EGLGTFE   V+ +   +GFDG+AV+LL  N+ +G++ AA
Sbjct: 236 NIMTSMLISGAFAGLAGAMEGLGTFEYAAVKGAFTGVGFDGIAVALLGGNTAVGVVLAAC 295

Query: 310 LFGVLSVGAPGMNI-AGIPPELIKVVTASIIFFVGVHYIIEYVIKPKKQ 357
           L G L +GA  M I +G+P E++ +V A II FV   Y I +V+   K+
Sbjct: 296 LLGGLKIGALNMPIESGVPSEVVDIVIAIIILFVASSYAIRFVMGKLKK 344
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2149> which encodes the amino acid sequence <SEQ ID 2150>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.74   Transmembrane  205 - 221 ( 200 - 228)
   INTEGRAL    Likelihood =-12.42   Transmembrane   21 -  37 (  14 -  45)
   INTEGRAL    Likelihood = -7.22   Transmembrane  115 - 131 ( 111 - 135)
   INTEGRAL    Likelihood = -4.78   Transmembrane  251 - 267 ( 249 - 269)
   INTEGRAL    Likelihood = -2.50   Transmembrane   70 -  86 (  69 -  86)
   INTEGRAL    Likelihood = -2.34   Transmembrane  302 - 318 ( 300 - 318)
   INTEGRAL    Likelihood = -1.44   Transmembrane  148 - 164 ( 147 - 165)
   INTEGRAL    Likelihood = -1.33   Transmembrane  326 - 342 ( 326 - 342)

----- Final Results -----
            bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 293/358 (81%), Positives = 333/358 (92%), Gaps = 1/358 (0%)

Query: 6    RRREMSKKAQKIAVPLISVVLGIILGAIIMLIFGYDPLWGYEGLFQTAFGSIKNIGEIFR 65
            RR+ MSK AQKIAVPLISV+LG +LGAIIM+IFGYDP+WGYEGLFQ AFGS+KNIGEIFR
Sbjct: 6    RRKVMSKNAQKIAVPLISVLLGFLLGAIIMVIFGYDPIWGYEGLFQIAFGSVKNIGEIFR 65

Query: 66   AMGPLILIALGFSVASRAGFFNIGLPGQALSGWIAAGWFALSHPDMPRPAMILCTIIIGI 125
            +MGPLILIALGF+VASRAGFFN+GL GQAL+GWI+AGWFAL +PDMPRP +IL T +IG+
Sbjct: 66   SMGPLILIALGFTVASRAGFFNVGLSGQALAGWISAGWFALLNPDMPRPLLILMTALIGM 125

Query: 126  VAGGITGAIPGILRAYLGTSEVIVTIMMNYIVLYSGNAIVQRVFPKSIMRTSDSSVYVSA 185
            +AGGI GAIPGILRAYLGTSEVIVTIMMNYI+LY GNAIVQR +P+S+ ++ DS++ VS
Sbjct: 126  IAGGIAGAIPGILRAYLGTSEVIVTIMMNYIILYVGNAIVQRGYPESVKQSIDSTIQVSD 185

Query: 186  NASYQTDWLSSLTNNSRINIGIFIAIIAVVLVWFLLNKTTLGFEIRSVGLNPNASEYAGM 245
            NASYQT WLS+LTNNSRINIGIF AIIA+ L+WFLLNKTTLGFEIRSVGLNP+ASEYAGM
Sbjct: 186  NASYQTHWLSALTNNSRINIGIFFAIIAIALIWFLLNKTTLGFEIRSVGLNPHASEYAGM 245

Query: 246  SAKRTIILSMIISGAFAGLGGVVEGLGTFENVFVQPSSLAIGFDGMAVSLLAANSPIGIL 305
            S+KRTIILSMIISGA AGLGGVVEGLGTFENVFVQ SSLA+GFDGMAVSLLAANSP+GI
Sbjct: 246  SSKRTIILSMIISGALAGLGGVVEGLGTFENVFVQGSSLAVGFDGMAVSLLAANSPLGIF 305

Query: 306  FAAFLFGVLSVGAPGMNIAGIPPELIKVVTASIIFFVGVHYIIE-YVIKPKKQMKGGK 362
```

```
           F++FLFGVL++GAPGMNIAGIPPEL+KVVTASIIFFVG HY+IE Y+I+PKK +KGGK
Sbjct: 306 FSSFLFGVLNIGAPGMNIAGIPPELVKVVTASIIFFVGSHYLIERYIIRPKKLVKGGK 363
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1520

A DNA sequence (GBSx1610) was identified in *S.agalactiae* <SEQ ID 4675> which encodes the amino acid sequence <SEQ ID 4676>. This protein is predicted to be sugar ABC transporter, ATP-binding protein (mglA). Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3851(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9803> which encodes amino acid sequence <SEQ ID 9804> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15144 GB:Z99120 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 311/497 (62%), Positives = 396/497 (79%), Gaps = 1/497 (0%)

Query: 14  VIEMKEITKKFGDFVANDHINLTVEKGEIHALLGENGAGKSTLMNMLAGLLEPTDGQIFI 73
            VIEM  I K F   VAND+INL V+KGEIHALLGENGAGKSTLMN+L GL +P  G+I +
Sbjct: 4   VIEMLNIRKAFPGIVANDNINLQVKKGEIHALLGENGAGKSTLMNVLFGLYQPERGEIRV 63

Query: 74  NGQPVTIDSPSKSSQLGIGMVHQHFMLVEAFTVAENIVLGNETTQNGVLDIKTAAKEIKE 133
             G+ V I+SP+K++ LGIGMVHQHFMLV+ FTVAENI+LG E  + G +D K A +E+++
Sbjct: 64  RGEKVHINSPNKANDLGIGMVHQHFMLVDTFTVAENIILGKEPKKFGRIDRKRAGQEVQD 123

Query: 134 LSEKYGLSVNPNAKISDISVGAQQRVEILKTLYRGADILIFDEPTAVLTPSEIKELMTIM 193
            +S++YGL ++P AK +DISVG QQR EILKTLYRGADILIFDEPTAVLTP EIKELM IM
Sbjct: 124 ISDRYGLQIHPEAKAADISVGMQQRAEILKTLYRGADILIFDEPTAVLTPHEIKELMQIM 183

Query: 194 KSLVKEGKSIILITHKLDEIRAVADKVTVIRRGKSIETVPVAGASSQQLAEMMVGRSVSF 253
            K+LVKEGKSIILITHKL EI  + D+VTVIR+GK I+T+ V   +  +LA +MVGR VSF
Sbjct: 184 KNLVKEGKSIILITHKLKEIMEICDRVTVIRKGKGIKTLDVRDTNQDELASLMVGREVSF 243

Query: 254 RTEKKEANPTDIILSVKDLVVEENRGGVLAVKNLSLDVRAGEIVGIAGIDGNGQSELIQA 313
            +TEK+ A P   +L++  + V++ R G+  V++LSL V+AGEIVGIAG+DGNGQSELI+A
Sbjct: 244 KTEKRAAQPGAEVLAIDGITVKDTR-GIETVRDLSLSVKAGEIVGIAGVDGNGQSELIEA 302

Query: 314 ITGLRKVTSGQIVIKGKDVTKFSSRQITELSVGHVPEDRHRDGLVLDMTMAENLALQTYY 373
            +TGLRK  SG I + GK +   + R+ITE  +GH+P+DRH+ GLVLD  + EN+ LQ+YY
Sbjct: 303 VTGLRKTDSGTITLNGKQIQNLTPRKITESGIGHIPQDRHKHGLVLDFPIGENILLQSYY 362

Query: 374 KEPLSHKGILNFAKIKEYARQLMTEFDVRGAGEHVLARGFSGGNQQKAIIAREVDRDPDL 433
            K+P S  G+L+  ++ + AR L+TE+DVR   E+  AR  SGGNQQKAII RE+DR+PDL
Sbjct: 363 KKPYSALGVLHKGEMYKKARSLITEYDVRTPDEYTHARALSGGNQQKAIIGREIDRNPDL 422

Query: 434 LIVSQPTRGLDVGAIEYIHKRLIEERDKGKAVLVVSFELDEILNLSDRIAVIHDGKIQGI 493
            LI +QPTRGLDVGAIE++HK+LIE+RD GKAVL++SFEL+EI+NLSDRIAVI +G+I
Sbjct: 423 LIAAQPTRGLDVGAIEFVHKKLIEQRDAGKAVLLLSFELEEIMNLSDRIAVIFEGRIIAS 482

Query: 494 VKPDQTNKQELGILMAG 510
            V P +T +QELG+LMAG
```

```
Sbjct: 483 VNPQETTEQELGLLMAG 499
 Identities = 75/242 (30%), Positives = 128/242 (51%), Gaps = 24/242 (9%)

Query: 280 GVLAVKNLSLDVRAGEIVGIAGIDGNGQSELIQAITGLRKVTSGQIVIKGKDVTKFSSRQ 339
           G++A  N++L V+ GEI  + G +G G+S L+  + GL +   G+I ++G+ V   S  +
Sbjct: 16  GIVANDNINLQVKKGEIHALLGENGAGKSTLMNVLFGLYQPERGEIRVRGEKVHINSPNK 75

Query: 340 ITELSVGHVPEDRHRDGLVLD-MTMAENLALQTYYKEPLSHKGILNFAKI--KEYARQLM 396
             +L +G V    H+  +++D  T+AEN+ L    KEP        F +I  K   +++
Sbjct: 76  ANDLGIGMV----HQHFMLVDTFTVAENIILG---KEPKK------FGRIDRKRAGQEVQ 122

Query: 397 TEFDVRGAGEHVLARG--FSGGNQQKAIIAREVDRDPDLLIVSQPTRGL---DVGAIEYI 451
              D  G   H  A+    S G QQ+A I + + R  D+LI  +PT  L   ++  +  I
Sbjct: 123 DISDRYGLQIHPEAKAADISVGMQQRAEILKTLYRGADILIFDEPTAVLTPHEIKELMQI 182

Query: 452 HKRLIEERDKGKAVLVVSFELDEILNLSDRIAVIHDGKIQGIVKPDQTNKQELGILMAGG 511
            K L++E   GK++++++ +L EI+ + DR+ VI  GK     +    TN+ EL  LM G
Sbjct: 183 MKNLVKE---GKSIILITHKLKEIMEICDRVTVIRKGKGIKTLDVRDTNQDELASLMVGR 239

Query: 512 KI 513
           ++
Sbjct: 240 EV 241
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4677> which encodes the amino acid sequence <SEQ ID 4678>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3558(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 431/511 (84%), Positives = 467/511 (91%), Gaps = 1/511 (0%)

Query: 10  MTQNVIEMKEITKKFGDFVANDHINLTVEKGEIHALLGENGAGKSTLMNMLAGLLEPTDG 69
           MTQ+VIEM+EITKKFGDFVANDHINL V KGEIHALLGENGAGKSTLMNMLAGLLEPT G
Sbjct: 7   MTQHVIEMREITKKFGDFVANDHINLNVRKGEIHALLGENGAGKSTLMNMLAGLLEPTSG 66

Query: 70  QIFINGQPVTIDSPSKSSQLGIGMVHQHFMLVEAFTVAENIVLGNETTQNGVLDIKTAAK 129
           +I IN +PV IDSPSKS++LGIGMVHQHFMLVEAFTVAENI+LGNE  +NG LD+  A+K
Sbjct: 67  EIVINDKPVQIDSPSKSAKLGIGMVHQHFMLVEAFTVAENIILGNEVVKNGCLDLNQASK 126

Query: 130 EIKELSEKYGLSVNPNAKISDISVGAQQRVEILKTLYRGADILIFDEPTAVLTPSEIKEL 189
           +IK LSEKYGL++NP+AK+SDISVGAQQRVEILKTLYRGADILIFDEPTAVLTP+EIKEL
Sbjct: 127 DIKVLSEKYGLAINPSAKVSDISVGAQQRVEILKTLYRGADILIFDEPTAVLTPAEIKEL 186

Query: 190 MTIMKSLVKEGKSIILITHKLDEIRAVADKVTVIRRGKSIETVPVAGASSQQLAEMMVGR 249
           MTIMK+LVKEGKSIILITHKLDEIRAVAD+VTVIRRGKSIETV VAGA+SQ LAEMMVGR
Sbjct: 187 MTIMKNLVKEGKSIILITHKLDEIRAVADRVTVIRRGKSIETVDVAGATSQDLAEMMVGR 246

Query: 250 SVSFRTEKKEANPTDIILSVKDLVVEENRGGVLAVKNLSLDVRAGEIVGIAGIDGNGQSE 309
           SVSF T KK A P D++LS+K+L V+ENR GV AVK LSLDVRAGEIVGIAGIDGNGQSE
Sbjct: 247 SVSFTTSKKAAEPKDVVLSIKNLEVDENR-GVPAVKGLSLDVRAGEIVGIAGIDGNGQSE 305

Query: 310 LIQAITGLRKVTSGQIVIKGKDVTKFSSRQITELSVGHVPEDRHRDGLVLDMTMAENLAL 369
           LIQAITGLRKV SG I+IK  +VT  SSR+ITELSVGHVPEDRHRDGL+LD+++AEN AL
Sbjct: 306 LIQAITGLRKVKSGSIMIKNNEVTHLSSRKITELSVGHVPEDRHRDGLILDLSLAENTAL 365

Query: 370 QTYYKEPLSHKGILNFAKIKEYARQLMTEFDVRGAGEHVLARGFSGGNQQKAIIAREVDR 429
           QTYYK+PLS  GILN+ KI +YARQLM EFDVRGA E V ARGFSGGNQQKAIIAREVDR
Sbjct: 366 QTYYKQPLSQNGILNYTKINDYARQLMKEFDVRGANELVPARGFSGGNQQKAIIAREVDR 425

Query: 430 DPDLLIVSQPTRGLDVGAIEYIHKRLIEERDKGKAVLVVSFELDEILNLSDRIAVIHDGK 489
```

```
           DPDLLIVSQPTRGLDVGAIEYIHKRLI+ERDKGKAVLVVSFELDEILNLSDRIAVIHDGK
Sbjct: 426 DPDLLIVSQPTRGLDVGAIEYIHKRLIKERDKGKAVLVVSFELDEILNLSDRIAVIHDGK 485

Query: 490 IQGIVKPDQTNKQELGILMAGGKIEKEERDV 520
           IQGIV P+ TNKQELGILMAGG I KEE  V
Sbjct: 486 IQGIVSPENTNKQELGILMAGGSIHKEEGHV 516
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1521

A DNA sequence (GBSx1612) was identified in *S.agalactiae* <SEQ ID 4679> which encodes the amino acid sequence <SEQ ID 4680>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
          [Bacillus subtilis]
 Identities = 164/335 (48%), Positives = 224/335 (65%), Gaps = 10/335 (2%)

Query: 18  LAACGHRGASKSGGKS-DSLKVAMVTDTGGVDDKSFNQSGWEGMQAWGKKNGLKKGA-GF 75
            L ACG+   S   G+   +  VAMVTD GGVDDKSFNQS WEG+QA+GK+NGLKKG  G+
Sbjct: 11  LGACGNSEKSSGSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAFGKENGLKKGKNGY 70

Query: 76  DYFQSASESDYATNLDTAVSSGYKLIFGIGFSLHDAIDKAADNNKDVNYVIVDDVIKGKD 135
            DY QS S++DY TNL+      + LI+G+G+ + D+I + AD  K+ N+ I+D V+  KD
Sbjct: 71  DYLQSKSDADYTTNLNKLARENFDLIYGVGYLMEDSISEIADQRKNTNFAIIDAVVD-KD 129

Query: 136 NVASVVFADNESAYLAGIAAAKTTKTKTVGFVGGMESEVITRFEKGFEAGVKSVDKSIKI 195
            NVAS+ F + E ++L G+AAA ++K+  +GFVGGMESE+I +FE GF AGV++V+    +
Sbjct: 130 NVASITFKEQEGSFLVGVAAALSSKSGKIGFVGGMESELIKKFEVGFRAGVQAVNPKAVV 189

Query: 196 KVDYAGSFGDAAKGKTIAAAQYASGADIVYQVAGGTGAGVFSEAKSRNESLKEADKVWVL 255
            +V YAG F  A  GK  A + Y SG D++Y  AG TG GVF+EAK+  +   + D VWV+
Sbjct: 190 EVKYAGGFDKADVGKATAESMYKSGVDVIYHSAGATGTGVFTEAKNLKKEDPKRD-VWVI 248

Query: 256 GVDRDQAAEGKYTSKDGKASNFVLASSIKEVGKSVELIATKTSKGKFPGGNVTTYGLKDG 315
            GVD+DQ AEG+    +G   N  L S +K+V   VE +  K S GKFPGG   TYGL
Sbjct: 249 GVDKDQYAEGQV---EGTDDNVTLTSMVKKVDTVVEDVTKKASDGKFPGGETLTYGLDQD 305

Query: 316 GVDIATT--NLSDDAVKAIKEAKAKIISGDIKVPS 348
            GV I+ +  NLSDD +KA+ + K KII G +++P+
Sbjct: 306 GVGISPSKQNLSDDVIKAVDKWKKKIIDG-LEIPA 339
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 861> which encodes the amino acid sequence <SEQ ID 862>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 275/351 (78%), Positives = 312/351 (88%), Gaps = 3/351 (0%)

Query: 1   MNKKIAGIGLASIAVLSLAACGHRGASKSG--GKSDSLKVAMVTDTGGVDDKSFNQSGWE 58
            MNKK  G+GLAS+AVLSLAACG+RGASK G  GK+D LKVAMVTDTGGVDDKSFNQS WE
Sbjct: 1   MNKKFIGLGLASVAVLSLAACGNRGASKGGASGKTD-LKVAMVTDTGGVDDKSFNQSAWE 59

Query: 59  GMQAWGKKNGLKKGAGFDYFQSASESDYATNLDTAVSSGYKLIFGIGFSLHDAIDKAADN 118
            G+Q+WGK+ GL+KG GFDYFQS SES+YATNLDTAVS GY+LI+GIGF+L DAI KAA +
Sbjct: 60  GLQSWGKEMGLQKGTGFDYFQSTSESEYATNLDTAVSGGYQLIYGIGFALKDAIAKAAGD 119

Query: 119 NKDVNYVIVDDVIKGKDNVASVVFADNESAYLAGIAAAKTTKTKTVGFVGGMESEVITRF 178
            N+ V +VI+DD+I+GKDNVASV FAD+E+AYLAGIAAAKTTKTKTVGFVGGME  VITRF
Sbjct: 120 NEGVKFVIIDDIIEGKDNVASVTFADHEAAYLAGIAAAKTTKTKTVGFVGGMEGTVITRF 179

Query: 179 EKGFEAGVKSVDKSIKIKVDYAGSFGDAAKGKTIAAAQYASGADIVYQVAGGTGAGVFSE 238
            EKGFEAGVKSVD +I++KVDYAGSFGDAAKGKTIAAAQYA+GAD++YQ AGGTGAGVF+E
Sbjct: 180 EKGFEAGVKSVDDTIQVKVDYAGSFGDAAKGKTIAAAQYAAGADVIYQAAGGTGAGVFNE 239

Query: 239 AKSRNESLKEADKVWVLGVDRDQAAEGKYTSKDGKASNFVLASSIKEVGKSVELIATKTS 298
            AK+ NE   EADKVWV+GVDRDQ  EGKYTSKDGK +NFVLASSIKEVGK+V+LI  + +
Sbjct: 240 AKAINEKRSEADKVWVIGVDRDQKDEGKYTSKDGKEANFVLASSIKEVGKAVQLINKQVA 299

Query: 299 KGKFPGGNVTTYGLKDGGVDIATTNLSDDAVKAIKEAKAKIISGDIKVPSK 349
              KFPGG  T YGLKDGGV+IATTN+S +AVKAIKEAKAKI SGDIKVP K
Sbjct: 300 DKKFPGGKTTVYGLKDGGVEIATTNVSKEAVKAIKEAKAKIKSGDIKVPEK 350
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9061> which encodes amino acid sequence <SEQ ID 9062>. Analysis of this protein sequence reveals the following:

```
     Possible site: 17

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  414 bits (1052), Expect = e-117
 Identities = 196/347 (56%), Positives = 253/347 (72%), Gaps = 2/347 (0%)

Query: 1   MNKKVMSLGLVSTALFTLGGCTNNSAKQT--TDNSLKIAMITNQTGIDDKSFNQSAWEGL 58
            MNKK+  +GL S A+ +L  C +  A ++    +SLK+AM+T+  G+DDKSFNQS WEG+
Sbjct: 1   MNKKIAGIGLASIAVLSLAACGHRGASKSGGKSDSLKVAMVTDTGGVDDKSFNQSGWEGM 60

Query: 59  QAWGKENKLEKGKGYDYFQSANESEFTTNLESAVTNGYNLVFGIGFPLHDAVEKVAANNP 118
            QAWGK+N L+KG G+DYFQSA+ES++ TNL++AV++GY L+FGIGF LHDA++K A NN
Sbjct: 61  QAWGKKNGLKKGAGFDYFQSASESDYATNLDTAVSSGYKLIFGIGFSLHDAIDKAADNNK 120

Query: 119 DNHFAIVDDVIKGQKNVASITFSDHEAAYLAGVXXXXXXXXXQVGFVGGMEGDVVKRFEK 178
            D ++ IVDDVIKG+ NVAS+ F+D+E+AYLAG+          VGFVGGME +V+ RFEK
Sbjct: 121 DVNYVIVDDVIKGKDNVASVVFADNESAYLAGIAAAKTTKTKTVGFVGGMESEVITRFEK 180

Query: 179 GFEAGVKSVDDTIKVRVAYAGSFXXXXXXXXXXXXXXXEGADVIYHAAGGTGAGVFSEAK 238
            GFEAGVKSVD +IK++V YAGSF                GAD++Y  AGGTGAGVFSEAK
Sbjct: 181 GFEAGVKSVDKSIKIKVDYAGSFGDAAKGKTIAAAQYASGADIVYQVAGGTGAGVFSEAK 240

Query: 239 SINEKRKEEDKVWVIGVDRDQSEDGKYTTKDGKSANFVLTSSIKEVGKALVKVAVKTSED 298
            S NE  KE DKVWV+GVDRDQ+ +GKYT+KDGK++NFVL SSIKEVGK++  +A KTS+
Sbjct: 241 SRNESLKEADKVWVLGVDRDQAAEGKYTSKDGKASNFVLASSIKEVGKSVELIATKTSKG 300

Query: 299 QFPGGQITTFGLKEGGVSLTTDALTQDTXXXXXXXXXXXXXGTITVP 345
            +FPGG +TT+GLK+GGV + T  L+ D              G I VP
Sbjct: 301 KFPGGNVTTYGLKDGGVDIATTNLSDDAVKAIKEAKAKIISGDIKVP 347
```

SEQ ID 4680 (GBS211) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 6; MW 40kDa).

The GBS211-His fusion product was purified (Figure 205, lane 8) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 259A) and FACS (Figure 259B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1522

A DNA sequence (GBSx1613) was identified in *S.agalactiae* <SEQ ID 4681> which encodes the amino acid sequence <SEQ ID 4682>. This protein is predicted to be cytidine deaminase (cdd). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2112(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9801> which encodes amino acid sequence <SEQ ID 9802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB51906 GB:AJ237978 cytidine deaminase [Bacillus psychrophilus]
 Identities = 66/114 (57%), Positives = 81/114 (70%)

Query: 26  KASENAYVPYSKFPVGAALRTAEGKIFTGCNVENISYGLANCAERTAIFKAVSEGYKDFS 85
            KA E AYVPYSKFPVGAAL   +G I+ GCN+EN +Y + NCAERTA FKAVS+G + F
Sbjct: 12  KAREQAYVPYSKFPVGAALLAEDGTIYHGCNIENSAYSMTNCAERTAFFKAVSDGVRSFK 71

Query: 86  EIAIYGNTEEPISPCGACRQVMVEFFNKNAKVTLIAKNGKTVETTVGELLPYSF 139
             +A+  +TE P+SPCGACRQV+ EF N +  V L    G   ETTV +LLP +F
Sbjct: 72  ALAVVADTEGPVSPCGACRQVIAEFCNGSMPVYLTNLKGDIEETTVAKLLPGAF 125
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4683> which encodes the amino acid sequence <SEQ ID 4684>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0041(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
           [Bacillus subtilis]
 Identities = 152/339 (44%), Positives = 223/339 (64%), Gaps = 11/339 (3%)

Query: 8   LGLVSTALFTLGGCTNN---SAKQTTDNSLKIAMITNQTGIDDKSFNQSAWEGLQAWGKE 64
            + LV  A   LG C N+   S      N   +AM+T+  G+DDKSFNQSAWEG+QA+GKE
Sbjct: 1   MSLVIAAGTILGACGNSEKSSGSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAFGKE 60
```

```
Query: 65  NKLEKGK-GYDYFQSANESEFTTNLESAVTNGYNLVFGIGFPLHDAVEKVAANNPDNHFA 123
           N L+KGK GYDY QS +++++TTNL       ++L++G+G+ + D++ ++A    + +FA
Sbjct: 61  NGLKKGKNGYDYLQSKSDADYTTNLNKLARENFDLIYGVGYLMEDSISEIADQRKNTNFA 120

Query: 124 IVDDVIKGQKNVASITFSDHEAAYLAGVAAAKTTKTKQVGFVGGMEGDVVKRFEKGFEAG 183
           I+D V+  + NVASITF + E ++L GVAAA ++K+ ++GFVGGME +++K+FE GF AG
Sbjct: 121 IIDAVVD-KDNVASITFKEQEGSFLVGVAAALSSKSGKIGFVGGMESELIKKFEVGFRAG 179

Query: 184 VKSVDDTIKVRVAYAGSFADAAKGKTIAAAQYAEGADVIYHAAGGTGAGVFSEAKSINEK 243
           V++V+    V V YAG F  A  GK  A + Y  G DVIYH+AG TG GVF+EAK++ ++
Sbjct: 180 VQAVNPKAVVEVKYAGGFDKADVGKATAESMYKSGVDVIYHSAGATGTGVFTEAKNLKKE 239

Query: 244 RKEEDKVWVIGVDRDQSEDGKYTTKDGKSANFVLTSSIKEVGKALVKVAVKTSEDQFPGG 303
             + D VWVIGVD+DQ  +G+    +G   N  LTS +K+V   +  V  K S+ +FPGG
Sbjct: 240 DPKRD-VWVIGVDKDQYAEGQV---EGTDDNVTLTSMVKKVDTVVEDVTKKASDGKFPGG 295

Query: 304 QITTFGLKEGGVSLTTDA--LTQDTKKAIEAAKKAIIEG 340
           +  T+GL + GV ++     L+ D  KA++  KK II+G
Sbjct: 296 ETLTYGLDQDGVGISPSKQNLSDDVIKAVDKWKKKIIDG 334
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 88/128 (68%), Positives = 107/128 (82%)

Query: 15  MGNIELKKLAVKASENAYVPYSKFPVGAALRTAEGKIFTGCNVENISYGLANCAERTAIF 74
           MG  +L   AV+ASE AYVPYS FPVGAAL+T +G I+TGCN+EN+S+GL NC ERTAIF
Sbjct: 1   MGTTDLVSCAVQASEYAYVPYSHFPVGAALKTKDGTIYTGCNIENVSFGLTNCGERTAIF 60

Query: 75  KAVSEGYKDFSEIAIYGNTEEPISPCGACRQVMVEFFNKNAKVTLIAKNGKTVETTVGEL 134
           KA+S+G+K+  EIAIYG T +P+SPCGACRQVM EFF+ ++ VTLIAKNG+TVE TVG+L
Sbjct: 61  KAISDGHKELVEIAIYGETMQPVSPCGACRQVMAEFFDPSSLVTLIAKNGQTVEMTVGDL 120

Query: 135 LPYSFVDL 142
           L YSF DL
Sbjct: 121 LLYSFTDL 128
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1523

A DNA sequence (GBSx1614) was identified in *S.agalactiae* <SEQ ID 4685> which encodes the amino acid sequence <SEQ ID 4686>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2979(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9799> which encodes amino acid sequence <SEQ ID 9800> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11882 GB:Z99104 alternate gene name: ybaA~similar to
           hypothetical proteins [Bacillus subtilis]
 Identities = 90/201 (44%), Positives = 144/201 (70%), Gaps = 5/201 (2%)

Query: 1   MANMYYTENPNVEHDIHELNVKLLGESFSFLTDAGVFSKRMIDYGSQVLLNSLHF-EKNK 59
           M+  YY+E P+V+ +    + +L  + F+F +D+GVFSK+ +D+GS++L++S    E
Sbjct: 1   MSEHYYSEKPSVKSNKQTWSFRLRNKDFTFTSDSGVFSKKEVDFGSRLLIDSFEEPEVEG 60
```

```
Query: 60  SLLDLGCGYGPLGISLAK-VQGVKATMVDINTRALELAKKNATRNGVV-VEVFQSNIYEN 117
            +LD+GCGYGP+G+SLA   +     M+D+N RA+EL+ +NA +NG+  V+++QS+++ N
Sbjct: 61  GILDVGCGYGPIGLSLASDFKDRTIHMIDVNERAVELSNENAEQNGITNVKIYQSDLFSN 120

Query: 118 I--SKTFDYIISNPPIRAGKQVVHSIIEESICYLNTGGSLTIVIQKKQGAPSAKAKMLDT 175
           +  ++TF  I++NPPIRAGK+VVH+I E+S  +L   G L IVIQKKQGAPSA  K+ +
Sbjct: 121 VDSAQTFASILTNPPIRAGKKVVHAIFEKSAEHLKASGELWIVIQKKQGAPSAIEKLEEL 180

Query: 176 FGNCDILKKDKGYYILRSEKV 196
           F    +++K KGYYI++++KV
Sbjct: 181 FDEVSVVQKKKGYYIIKAKKV 201
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4687> which encodes the amino acid sequence <SEQ ID 4688>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4232(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 139/195 (71%), Positives = 165/195 (84%)

Query: 1   MANMYYTENPNVEHDIHELNVKLLGESFSFLTDAGVFSKRMIDYGSQVLLNSLHFEKNKS 60
           M  MYY ENP+  HDIHE+ V+LL   F+FLTD+GVFSK+M+D+GSQVLL +L+F++N+
Sbjct: 12  MTKMYYDENPDSLHDIHEVKVELLNHPFTFLTDSGVFSKKMVDFGSQVLLKTLNFKENER 71

Query: 61  LLDLGCGYGPLGISLAKVQGVKATMVDINTRALELAKKNATRNGVVVEVFQSNIYENISK 120
           +LDLGCGYGPLGISLAKVQ V AT+VDIN RAL+LA+KNAT N V V +FQSNIYENIS
Sbjct: 72  VLDLGCGYGPLGISLAKVQRVDATLVDINNRALDLARKNATNNQVAVTIFQSNIYENISG 131

Query: 121 TFDYIISNPPIRAGKQVVHSIIEESICYLNTGGSLTIVIQKKQGAPSAKAKMLDTFGNCD 180
            F++IISNPPIRAGK+VVHSIIE+SI +L   G LTIVIQKKQGAPSAKAKM   FGN +
Sbjct: 132 HFEHIISNPPIRAGKRVVHSIIEKSIDFLVVNGDLTIVIQKKQGAPSAKAKMATIFGNVE 191

Query: 181 ILKKDKGYYILRSEK 195
           IL+KDKGYY+LRS K
Sbjct: 192 ILRKDKGYYVLRSIK 206
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1524

A DNA sequence (GBSx1615) was identified in *S.agalactiae* <SEQ ID 4689> which encodes the amino acid sequence <SEQ ID 4690>. This protein is predicted to be pantothenate kinase (coaA). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5021(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06594 GB:AP001516 pantothenate kinase [Bacillus halodurans]
 Identities = 140/307 (45%), Positives = 203/307 (65%), Gaps = 5/307 (1%)
```

```
Query: 4   EFINFDRISRENWKDLHQQSQALLTEKELESIKSLNDNINIQDVIDIYLPLINLIQIYKR 63
           +F  +  +SR  WK L + S   + E+ELE +  LN+ I + +V DIY+PL  L+ ++
Sbjct: 8   DFFPYTVLSRSQWKSLRKASSLPINEQELEQLVGLNEPITLNEVADIYVPLAELLHVHAT 67

Query: 64  SQENLSFSKAIFLKKENYQRPFIIGISGSVAVGKSTTSRLLQLLISRTFKDSHVELVTTD 123
           + + L   K  F      + PFIIG++GSVAVGKSTT+RLLQ L+    +  HV+LVTTD
Sbjct: 68  AYQRLQQQKRGFFHHGKNRSPFIIGLAGSVAVGKSTTARLLQKLLKAWPEHHHVDLVTTD 127

Query: 124 GFLYPNEKLIQNGILNRKGFPESYDMESLLNFLDTIKNGIT-AKIPIYSHEIYDIVPNQL 182
           GFLYPNE L   G++++KGFPESYD+ +L+ FL  +K G    K P+YSH  Y+IV
Sbjct: 128 GFLYPNETLEARGLMDKKGFPESYDLPALIRFLSDVKAGEPYVKAPVYSHLTYNIVEGDY 187

Query: 183 QTIETPDFLILEGINVFQ-NQQNHRL---YMNDYFDFSIYIDAENKQIEEWYLQRFNSLL 238
           Q +  PD +I+EGINV Q N++NH +   +++D+FDFSIY+DA+ +QI +WY++RF  L
Sbjct: 188 QVVHEPDIVIVEGINVLQVNKRNHHIPNVFVSDFFDFSIYVDAKEEQILQWYIERFKLLQ 247

Query: 239 QLAEADPSNYYHKFTQIPPHKAMELAKDIWKTINLVNLEKYIEPTRNRADFIIHKGKHHK 298
             A  DP++Y+H+F  +   +A + A  IWK IN VNL + I PT++RAD ++ KG HH
Sbjct: 248 NTAFQDPNSYFHRFRHLSEVEAEQFATSIWKNINGVNLHENILPTKHRADLVLQKGPHHF 307

Query: 299 IDEIYLK 305
           IDE+ L+
Sbjct: 308 IDEVKLR 314
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4691> which encodes the amino acid sequence <SEQ ID 4692>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
           bacterial cytoplasm --- Certainty=0.4790(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 219/306 (71%), Positives = 269/306 (87%)

Query: 1   MNNEFINFDRISRENWKDLHQQSQALLTEKELESIKSLNDNINIQDVIDIYLPLINLIQI 60
           M+NEFINF++ISRE+WK LHQ+++ALLT++EL+SI SLNDNI+I DVIDIYLPLINLIQ+
Sbjct: 1   MSNEFINFEKISRESWKTLHQKAKALLTQEELKSITSLNDNISINDVIDIYLPLINLIQV 60

Query: 61  YKRSQENLSFSKAIFLKKENYQRPFIIGISGSVAVGKSTTSRLLQLLISRTFKDSHVELV 120
           YK +QENLSFSK++FLKK+   RPFIIGISGSVAVGKSTTSRLLQLL+SRT  +S VELV
Sbjct: 61  YKIAQENLSFSKSLFLKKDIQLRPFIIGISGSVAVGKSTTSRLLQLLLSRTHPNSQVELV 120

Query: 121 TTDGFLYPNEKLIQNGILNRKGFPESYDMESLLNFLDTIKNGITAKIPIYSHEIYDIVPN 180
           TTDGFLYPN+ LI+ G+LNRKGFPESY+ME LL+FLD+IKNG TA  P+YSH+IYDI+PN
Sbjct: 121 TTDGFLYPNQFLIEQGLLNRKGFPESYNMELLLDFLDSIKNGQTAFAPVYSHDIYDIIPN 180

Query: 181 QLQTIETPDFLILEGINVFQNQQNHRLYMNDYFDFSIYIDAENKQIEEWYLQRFNSLLQL 240
           Q Q+   PDFLI+EGINVFQNQQN+RLYM+DYFDFSIYIDA++  IE WY++RF S+L+L
Sbjct: 181 QKQSFNNPDFLIVEGINVFQNQQNNRLYMSDYFDFSIYIDADSSHIETWYIERFLSILKL 240

Query: 241 AEADPSNYYHKFTQIPPHKAMELAKDIWKTINLVNLEKYIEPTRNRADFIIHKGKHHKID 300
           A+ DP NYY ++ Q+P  +A+  A+++WKT+NL NLEK+IEPTRNRA+ I+HK   HKID
Sbjct: 241 AKRDPHNYYAQYAQLPRSEAIAFARNVWKTVNLENLEKFIEPTRNRAELILHKSADHKID 300

Query: 301 EIYLKK 306
           EIYLKK
Sbjct: 301 EIYLKK 306
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1525

A DNA sequence (GBSx1616) was identified in *S.agalactiae* <SEQ ID 4693> which encodes the amino acid sequence <SEQ ID 4694>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3866(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05058 GB:AP001511 ribosomal protein S20 (BS20) [Bacillus halodurans]
 Identities = 47/86 (54%), Positives = 59/86 (67%), Gaps = 7/86 (8%)

Query: 3  VKTLANIKSAIKRAELNVKQNEKNSAQKSAMRTAIKAFEA---NPSEELYRA----ASSS 55
           +K  ANIKSAIKR + N K+  +N++ KSA+RTAIK FEA   N   E  +A    A+
Sbjct: 1  MKGNANIKSAIKRVKTNEKRRIQNASVKSALRTAIKQFEAKVENNDAEAAKAAFVEATKK 60

Query: 56 IDKAASKGLIHTNKASRDKARLATKL 81
           +DKAA+KGLIH N ASR K+RLA KL
Sbjct: 61 LDKAANKGLIHKNAASRQKSRLAKKL 86
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4695> which encodes the amino acid sequence <SEQ ID 4696>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3872(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 76/82 (92%), Positives = 78/82 (94%)

Query: 1  MEVKTLANIKSAIKRAELNVKQNEKNSAQKSAMRTAIKAFEANPSEELYRAASSSIDKAA 60
           +EVKTLANIKSAIKRAELNVK NEKNSAQKSAMRTAIKAFEANPSEEL+RAASSSIDKA
Sbjct: 1  LEVKTLANIKSAIKRAELNVKANEKNSAQKSAMRTAIKAFEANPSEELFRAASSSIDKAE 60

Query: 61 SKGLIHTNKASRDKARLATKLG 82
           SKGLIH NKASRDKARLA KLG
Sbjct: 61 SKGLIHKNKASRDKARLAAKLG 82
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1526

A DNA sequence (GBSx1617) was identified in *S.agalactiae* <SEQ ID 4697> which encodes the amino acid sequence <SEQ ID 4698>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.99   Transmembrane   31 -  47 (  25 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC35851 GB:AF086736 amino acid-binding protein Abp
           [Streptococcus uberis]
 Identities = 169/269 (62%), Positives = 203/269 (74%), Gaps = 2/269 (0%)

Query: 29  KNILLTIIFGLFMIILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGD 88
            K ILLT +     + L ACG S+   A  D W+ Y+KEK IT+GFDNTFVPMGF+  SG
Sbjct: 4   KKILLTTLALASTLFLVACGKSSA--AKTDQWDTYKKEKSITLGFDNTFVPMGFKDESGK 61

Query: 89  YTGFDIDLANAVFKEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFTN 148
             TGFD++LA AVF+EYGI VK+QPINWD+KETEL NG ID+IWNGYS T ER  KVAF+
Sbjct: 62  NTGFDVELAKAVFQEYGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAFST 121

Query: 149 PYMNNHQVIVTKTSSHINSIKDMKGKKLGAQSGSSGFDAFNAKPDILKKFVKGKEAVQYD 208
            PYM N QV+VTK SS+I S   MKGK LGAQSGSSG+DAF + P +LK  VK  +A QY+
Sbjct: 122 PYMKNEQVLVTKKSSNITSFAAMKGKVLGAQSGSSGYDAFTSNPKVLKDIVKDNDATQYE 181

Query: 209 TFTQALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKTAYQGENFVVGARKVDRRL 268
            TF QA IDLKN+RIDGLLID+VYANYYLKQEG +  Y  VK+ + GE+F VG RK D+ L
Sbjct: 182 TFIQAFIDLKNDRIDGLLIDKVYANYYLKQEGELTNYNIVKSEFDGEDFAVGVRKEDKIL 241

Query: 269 IEKINKAFKQLHNKGRFQKISYKWFGEDV 297
            ++ IN AF +L+  G+FQ+IS KWFGEDV
Sbjct: 242 LKNINSAFTKLYKTGKFQEISQKWFGEDV 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4699> which encodes the amino acid sequence <SEQ ID 4700>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21

>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC35851 GB:AF086736 amino acid-binding protein Abp
           [Streptococcus uberis]
 Identities = 176/277 (63%), Positives = 220/277 (78%), Gaps = 1/277 (0%)

Query: 1   MIIKKRTVAILAIASSFFLVACQATKSLKSGDAWGVYQKQKSITVGFDNTFVPMGYKDES 60
            M +KK  +  LA+AS+ FLVAC  + + K+ D W  Y+K+KSIT+GFDNTFVPMG+KDES
Sbjct: 1   MNLKKILLTTLALASTLFLVACGKSSAAKT-DQWDTYKKEKSITLGFDNTFVPMGFKDES 59

Query: 61  GRCKGFDIDLAKEVFHQYGLKVNFQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAF 120
            G+  GFD++LAK VF +YG+KV FQ INWD+KE EL NGKID+IWNGYS+TKERQ KVAF
Sbjct: 60  GKNTGFDVELAKAVFQEYGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAF 119

Query: 121 TDSYMRNEQIIVVKKRSDIKTISDMKHKVLGAQSASSGYDSLLRTPKLLKDFIKNKDANQ 180
            +  YM+NEQ++V KK S+I + + MK KVLGAQS SSGYD+    PK+LKD +K+ DA Q
Sbjct: 120 STPYMKNEQVLVTKKSSNITSFAAMKGKVLGAQSGSSGYDAFTSNPKVLKDIVKDNDATQ 179

Query: 181 YETFTQAFIDLKSDRIDGILIDKVYANYYLAKEGQLENYRMIPTTFENEAFSVGLRKEDK 240
            YETF QAFIDLK+DRIDG+LIDKVYANYYL +EG+L NY ++ + F+ E F+VG+RKEDK
Sbjct: 180 YETFIQAFIDLKNDRIDGLLIDKVYANYYLKQEGELTNYNIVKSEFDGEDFAVGVRKEDK 239

Query: 241 TLQAKINRAFRVLYQNGKFQAISEKWFGDDVATANIK 277
             L   IN AF  LY+ GKFQ IS+KWFG+DVAT N+K
Sbjct: 240 ILLKNINSAFTKLYKTGKFQEISQKWFGEDVATENVK 276
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 151/266 (56%), Positives = 189/266 (70%), Gaps = 4/266 (1%)

Query: 32  LLTIIFGLFMIILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGDYTG 91
            +L I    F++   AC  + K +   D W  YQK+K IT+GFDNTFVPMG++  SG   G
Sbjct: 10  ILAIASSFFLV---AC-QATKSLKSGDAWGVYQKQKSITVGFDNTFVPMGYKDESGRCKG 65

Query: 92  FDIDLANAVFKEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFTNPYM 151
            FDIDLA  VF +YG+ V +Q INWDMKE ELNNG ID+IWNGYS T ER  KVAFT+ YM
Sbjct: 66  FDIDLAKEVFHQYGLKVNFQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYM 125

Query: 152 NNHQVIVTKTSSHINSIKDMKGKKLGAQSGSSGFDAFNAKPDILKKFVKGKEAVQYDTFT 211
             N Q+IV K  S I +I DMK K LGAQS SSG+D+    P +LK F+K K+A QY+TFT
Sbjct: 126 RNEQIIVVKKRSDIKTISDMKHKVLGAQSASSGYDSLLRTPKLLKDFIKNKDANQYETFT 185

Query: 212 QALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKTAYQGENFVVGARKVDRRLIEK 271
            QA IDLK++RIDG+LID+VYANYYL +EG ++ Y  + T ++ E F VG RK D+ L  K
Sbjct: 186 QAFIDLKSDRIDGILIDKVYANYYLAKEGQLENYRMIPTTFENEAFSVGLRKEDKTLQAK 245

Query: 272 INKAFKQLHNKGRFQKISYKWFGEDV 297
            IN+AF+ L+  G+FQ IS KWFG+DV
Sbjct: 246 INRAFRVLYQNGKFQAISEKWFGDDV 271
```

A related GBS gene <SEQ ID 8833> and protein <SEQ ID 8834> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: 22   Crend: 4
       Sequence Pattern: CGMS
SRCFLG: 0
McG: Length  of  UR:   22
     Peak Value of UR:   3.05
     Net Charge of CR: 2
McG: Discrim Score:     11.16
GvH: Signal Score (-7.5): -1.96
     Possible site: 24
>>> May be a lipoprotein
Amino Acid Composition: calculated from 23
ALOM program   count: 0 value:   8.96 threshold:  0.0
   PERIPHERAL  Likelihood =  8.96      68
 modified ALOM score:  -2.29

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
62.2/75.8% over 270aa
                                                    Streptococcus uberis
  GP|3603430| amino acid-binding protein Abp Insert characterized

ORF00904(385 - 1203 of 1503)
GP|3603430|gb|AAC35851.1||AF086736(4 - 274 of 277) amino acid-binding protein Abp
{Streptococcus uberis}
%Match = 34.8
%Identity = 62.1  %Similarity = 75.7
Matches = 169  Mismatches = 65  Conservative Sub.s = 37

153       183       213       243       273       303       333       363
FHYLGGKSNVSH*LVR**LIHRLLVMMSQLALLIQSCVKK*KN*FYKIEKQV*HKL**HMIFNLLKVYLIRFSKLILSRL

393       423       453       483       513       543       573       603
GGRLLTHKNILLTIIFGLFMIILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGDYTGFDIDLANAVF
```

```
    :  | |||| :     : | ||| |:   |  | |: |:||| ||:|||||||||||:  ||  ||||::|| |||
    MNLKKILLTTLALASTLFLVACGKSS--AAKTDQWDTYKKEKSITLGFDNTFVPMGFKDESGKNTGFDVELAKAVF
            10        20          30        40        50        60        70

633       663       693       723       753       783       813       843
KEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFTNPYMNNHQVIVTKTSSHINSIKDMKGKKLGAQSG
:|||| ||:||||||:||||| || ||:|||||| | ||  ||||: ||| | ||:||| ||:| |   |||| ||||||
QEYGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAFSTPYMKNEQVLVTKKSSNITSFAAMKGKVLGAQSG
              90       100       110       120       130       140       150

873       903       933       963       993      1023      1053      1083
SSGFDAFNAKPDILKKFVKGKEAVQYDTFTQALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKTAYQGENFVVGA
|||:||| : | :||  ||  :| ||:|| ||:|||||:||||||||:||||||||||| :  |  ||: : ||:| ||
SSGYDAFTSNPKVLKDIVKDNDATQYETFIQAFIDLKNDRIDGLLIDKVYANYYLKQEGELTNYNIVKSEFDGEDFAVGV
             170       180       190       200       210       220       230

1113      1143      1173      1203      1233      1263      1293      1323
RKVDRRLIEKINKAFKQLHNKGRFQKISYKWFGEDVYSKE*KTRNFS*SFILRKN*IKNIDISDVF*VN*PSLVARRALS
|| |: |:: || || :|:  |:||:|| ||||||| ::  |
RKEDKILLKNINSAFTKLYKTGKFQEISQKWFGEDVATENVKK
             250       260       270
```

SEQ ID 8834 (GBS225) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 10; MW 32kDa). The GBS225-His fusion product was purified (Figure 205, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 266), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1527

A DNA sequence (GBSx1618) was identified in *S.agalactiae* <SEQ ID 4701> which encodes the amino acid sequence <SEQ ID 4702>. This protein is predicted to be arginine ABC transporter, ATP-binding protein (glnQ). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3229(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB49429 GB:U73111 glutamine transport ATP-binding protein GLNQ
          [Salmonella typhimurium]
 Identities = 94/210 (44%), Positives = 146/210 (68%), Gaps = 3/210 (1%)

Query: 1   MLELKNISKCYGQKEIFKDFNLTVEEGKILSLVGPSGGGKTTLLRMLAGLEKIDSGTIVH 60
            M+E KN+SK +G  ++  + +L + +G+++ ++GPSG GK+TLLR +  LE+I SG ++
Sbjct: 1   MIEFKNVSKHFGPTQVLHNIDLNIRQGEVVVIIGPSGSGKSTLLRCINKLEEITSGDLIV 60

Query: 61  DGKEVS---VDHLETLNLLGFVFQDFQLFPHLTVLDNLILSPVKTMGLSKELAKEKALVL 117
            DG +V+   VD        G VFQ F LFPHLT L+N++  P++  G+ KE A+++A  L
Sbjct: 61  DGLKVNDPKVDERLIRQEAGMVFQQFYLFPHLTALENVMFGPLRVRGVKKEEAEKQAKAL 120

Query: 118 LERLGLKDHALVYPFSLSGGQKQRVALARAMMIDPQIIGYDEPTSALDPELRQEVEKLIL 177
            L ++GL + A  YP  LSGGQ+QRVA+ARA+ + P+++ +DEPTSALDPELR EV K++
Sbjct: 121 LAKVGLAERAHHYPSELSGGQQQRVAIARALAVKPKMMLFDEPTSALDPELRHEVLKVMQ 180

Query: 178 QNRETGMTQIVVTHDLQFAESISDTILKIN 207
```

```
                   E GMT ++VTH++ FAE ++  ++ I+
Sbjct: 181 DLAEEGMTMVIVTHEIGFAEKVASRLIFID 210
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4703> which encodes the amino acid sequence <SEQ ID 4704>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2146(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 164/209 (78%), Positives = 183/209 (87%)

Query: 1   MLELKNISKCYGQKEIFKDFNLTVEEGKILSLVGPSGGGKTTLLRMLAGLEKIDSGTIVH 60
            MLELKNISK +GQK IF  FNLTV++G++LSLVGPS GGKTTLLRMLAGLE IDSG + +
Sbjct: 1   MLELKNISKQFGQKTIFDGFNLTVQDGEVLSLVGPSSGGKTTLLRMLAGLESIDSGQVFY 60

Query: 61  DGKEVSVDHLETLNLLGFVFQDFQLFPHLTVLDNLILSPVKTMGLSKELAKEKALVLLER 120
            +G++V +DHLE  NLLGFVFQDFQLFPHLTVLDNL LSP  TMG  K  AKEKAL LL R
Sbjct: 61  NGEDVGIDHLENRNLLGFVFQDFQLFPHLTVLDNLTLSPTITMGKQKADAKEKALDLLAR 120

Query: 121 LGLKDHALVYPFSLSGGQKQRVALARAMMIDPQIIGYDEPTSALDPELRQEVEKLILQNR 180
            LGLK+HA VYP+SLSGGQKQRVALARAMMIDPQIIGYDEPTSALDPELRQ VE LI+QNR
Sbjct: 121 LGLKEHAQVYPYSLSGGQKQRVALARAMMIDPQIIGYDEPTSALDPELRQTVEALIVQNR 180

Query: 181 ETGMTQIVVTHDLQFAESISDTILKINPK 209
            E G+TQIVVTHDL FAE+ISD I+++NPK
Sbjct: 181 EMGITQIVVTHDLVFAEAISDRIIRVNPK 209
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1528

A DNA sequence (GBSx1619) was identified in *S.agalactiae* <SEQ ID 4705> which encodes the amino acid sequence <SEQ ID 4706>. This protein is predicted to be amino acid ABC transporter, permease protein (glnP). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.12   Transmembrane  102 - 118 (  96 - 120)

----- Final Results -----
               bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9341> which encodes amino acid sequence <SEQ ID 9342> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA98402 GB:AP002545 ABC amino acid transporter permease
           [Chlamydophila pneumoniae J138]
 Identities = 55/127 (43%), Positives = 83/127 (65%), Gaps = 1/127 (0%)

Query: 3   AAIIAFTMNYAAYFAEIFRGGIESIPKGQYEAAKVLKFSKFQTVWYIVLPQVFKIVLPSV 62
            A IIA +MN AAY AE  RGGI S+  GQ+E+A VL + K+Q   YI+ PQVFK +LPS+
```

```
Sbjct: 89  AGIIALSMNSAAYLAENIRGGINSLSIGQWESAMVLGYKKYQIFVYIIYPQVFKNILPSL 148

Query: 63  FNETITLVKDSSLVYILGVGDLLLESKTAANRDATLAPMF-IAGGIYLLLIGLLTILSKQ 121
             NE ++L+K+SS++ ++GV +L   +K   +R+     M+ I  G+Y L+    + +S+
Sbjct: 149 TNEFVSLIKESSILMVVGVPELTKVTKDIVSRELNPMEMYLICAGLYFLMTSSFSCISRL 208

Query: 122 VEKRFNY 128
             EKR +Y
Sbjct: 209 SEKRRSY 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4707> which encodes the amino acid sequence <SEQ ID 4708>. Analysis of this protein sequence reveals the following:

```
        Possible site: 34
  >>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood =-11.57   Transmembrane   21 -  37 (   7 -  44)
     INTEGRAL    Likelihood =-10.93   Transmembrane  185 - 201 ( 178 - 206)
     INTEGRAL    Likelihood = -3.29   Transmembrane   63 -  79 (  62 -  81)

  ----- Final Results -----
                bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05181 GB:AP001512 ABC transporter (permease) [Bacillus halodurans]
 Identities = 86/206 (41%), Positives = 126/206 (60%), Gaps = 1/206 (0%)

Query: 4   IQQVLPSLLDGALVTLQVFFIVIILSIPLGAILAFLMKIPFKPLQWFLTLYVWMMRGTPL 63
            IQ  +P +L+G  VTLQ   + ++  + LG +LA      ++  +WF   Y  + RGTPL
Sbjct: 8   IQPFMPFMLEGVWVTLQFVSVSLLFGLVLGIVLAIFKISKYRLFRWFADFYTSIFRGTPL 67

Query: 64  LLQLIFFYYVLPSVGISFDRMPAAILAFTLNYAAYFAEIFRGGIEAIPKGQYEAAKVLKL 123
            +LQL+  Y  LP  G+   +  AA LAF LN AAY +EI R GI+A+ KGQ EAA+ L +
Sbjct: 68  ILQLLMIYLALPQFGVDISQFQAAFLAFGLNSAAYVSEIIRAGIQAVDKGQREAAEALGI 127

Query: 124 KPLQTIRYIILPQVFKIVLPSVFNEVINLVKDSSLVYVLGVGDLL-LASKTAANRDATLA 182
                 +  IILPQ  + +LP++FNE INL K+S++V V+GV DL+  A  T+A      L
Sbjct: 128 PYRPMMLRIILPQAMRNILPALFNEFINLTKESAIVSVIGVTDLMRRAQITSAETYLYLE 187

Query: 183 PMFIAGLIYLLLIGLVTIISKQVEKR 208
            P+   GLIY +L+  +T+I + +E+R
Sbjct: 188 PLLFVGLIYYVLVMGLTVIGRLLERR 213
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 112/130 (86%), Positives = 121/130 (92%)

Query: 1   MPAAIIAFTMNYAAYFAEIFRGGIESIPKGQYEAAKVLKFSKFQTVWYIVLPQVFKIVLP 60
            MPAAI+AFT+NYAAYFAEIFRGGIE+IPKGQYEAAKVLK    QT+ YI+LPQVFKIVLP
Sbjct: 84  MPAAILAFTLNYAAYFAEIFRGGIEAIPKGQYEAAKVLKLKPLQTIRYIILPQVFKIVLP 143

Query: 61  SVFNETITLVKDSSLVYILGVGDLLLESKTAANRDATLAPMFIAGGIYLLLIGLLTILSK 120
            SVFNE I LVKDSSLVY+LGVGDLLL SKTAANRDATLAPMFIAG IYLLLIGL+TI+SK
Sbjct: 144 SVFNEVINLVKDSSLVYVLGVGDLLLASKTAANRDATLAPMFIAGLIYLLLIGLVTIISK 203

Query: 121 QVEKRFNYYK 130
            QVEKRFNYY+
Sbjct: 204 QVEKRFNYYQ 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1529

A DNA sequence (GBSx1620) was identified in *S.agalactiae* <SEQ ID 4709> which encodes the amino acid sequence <SEQ ID 4710>. This protein is predicted to be minidiscs. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.66   Transmembrane   44 -  60 (  39 -  66)
   INTEGRAL    Likelihood = -7.96   Transmembrane  129 - 145 ( 123 - 147)
   INTEGRAL    Likelihood = -5.15   Transmembrane   13 -  29 (   9 -  33)
   INTEGRAL    Likelihood = -2.39   Transmembrane   94 - 110 (  94 - 110)

----- Final Results -----
               bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF49688 GB:AE003532 mnd gene product [Drosophila melanogaster]
 Identities = 48/145 (33%), Positives = 78/145 (53%), Gaps = 8/145 (5%)

Query: 7   IKQTYGLMTTIAMIVGVVIGSGIYFKVDDILKFTGGDVFLGMVILVLGSFSIVFGSLSIS 66
           +K+  GL+  +A+IVGV++GSGI+     +LKF+ G +   +++ VL     + G+L  +
Sbjct: 39  LKKQIGLLDGVAIIVGVIVGSGIFVSPKGVLKFS-GSIGQSLIVWVLSGVLSMVGALCYA 97

Query: 67  ELAIRTSESGGIFSYYEKYVSPALAATLGLFASFLYL-PTLTAIVSWVAAFYTLGE---- 121
           EL     +SGG ++Y      P L A L L+ + L L PT  AI +   A Y L
Sbjct: 98  ELGTMIPKSGGDYAYIGTAFGP-LPAFLYLWVALLILVPTGNAITALTFAIYLLKPFWPS 156

Query: 122 -SSSLESQIILAAVYILALSLMNIF 145
              + +E+  +LAA  I  L+L+N +
Sbjct: 157 CDAPIEAVQLLAAAMICVLTLINCY 181
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1530

A DNA sequence (GBSx1621) was identified in *S.agalactiae* <SEQ ID 4711> which encodes the amino acid sequence <SEQ ID 4712>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1531

A DNA sequence (GBSx1622) was identified in *S.agalactiae* <SEQ ID 4713> which encodes the amino acid sequence <SEQ ID 4714>. This protein is predicted to be TRK potassium uptake system protein. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane  232 - 248 ( 232 - 248)

----- Final Results -----
               bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8835> which encodes amino acid sequence <SEQ ID 8836> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:     -4.65
GvH: Signal Score (-7.5): -3.64
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -0.06 threshold:  0.0
   INTEGRAL    Likelihood = -0.06   Transmembrane  228 - 244 ( 228 - 244)
   PERIPHERAL  Likelihood =  1.27     428
 modified ALOM score:   0.51

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB90401 GB:AE001046 TRK potassium uptake system protein
           (trkA-2) [Archaeoglobus fulgidus]
 Identities = 136/446 (30%), Positives = 238/446 (52%), Gaps = 12/446 (2%)

Query: 5   MRIIVVGGGKVGTALCRSLVAEKHDVVLIEKKENVLKRVTKQHDIMGIVGNGANYKILEQ 64
           MRI++ G G+VG  L  SL A  HDV++IEK  +  +RV++  D++ I GN AN K+L
Sbjct: 1   MRIVIAGAGEVGYHLAMSL-APNHDVIIIEKDVSRFERVSEL-DVVAINGNAANMKVLRD 58

Query: 65  AEVKNCDIFIAITDRDEVNMISAVLAKKMGAKETVVRMRNPEYSNPYFKDKNFLGFSSVV 124
           A V+  D+F+A+T  DEVN++S + AKK+GAK  +VR+ NPEY +     ++ LG+  ++
Sbjct: 59  AGVERADVFLAVTGNDEVNLLSGLAAKKVGAKNVIVRVENPEYVDRPIVKEHPLGYDVLI 118

Query: 125 NPELLAAQYIANTIEFPNATSVEHFANGRVMLMEFKILEGNKLCHTSMSQIRKKFGNIVI 184
            P+L  AQ  A  I  P A  V  F+ G+V ++E +++EG+K    +++ +     N+VI
Sbjct: 119 CPQLSLAQEAARLIGIPGAIEVVTFSGGKVEMIELQVMEGSKADGKAIADLYLP-QNVVI 177

Query: 185 CAIERDGKLIIPDGDATIQVKDKIFVTGNRIEMILFHNYVKNKVVKNLMVIGAGRIAYYL 244
            +I R+G + IP GD  ++  D++ +     ++ +        V + + + GAG I  Y
Sbjct: 178 ASIYRNGHIEIPRGDTVLRAGDRVAIVSKTEDVEMLKGIFGPPVTRRVTIFGAGTIGSYT 237

Query: 245 LNILKNTNTHVKLVELNQEQAEYFSQEFPNVPVVHGDGTAKNILLEESVTSFDAVATLTG 304
             IL    T VKL+E + E+ E  S E   V +V GD T    L+EE +   DAV   T
Sbjct: 238 AKILAKGMTSVKLIESSMERCEALSGELEGVRIVCGDATDIEFLIEEEIGKSDAVLAATE 297

Query: 305 VDEENIITSMFLESIGIPKNITKVNRTSLLEIIDDKQLSSIITPKRIAVDHVMHFVRGRV 364
            DE+N++ S+  +++G    I KV +   +++ +   +   + P+ +  + V   +R
Sbjct: 298 SDEKNLLISLLSKNLGARIAIAKVEKREYVKLFEAVGVDVALNPRSVTYNEVSKLLR--- 354

Query: 365 NAQDSNLEAMHHIANDRIETLQFEIKETSKLANRSLASLKLKQNILIAAIIRNNKTIFPT 424
```

```
                 +E +  I    +  +   +   ++L  ++L  L L ++ +I AI+R N+ + P
Sbjct: 355 ---TMRIETLAEIEGTAVVEV---VVRNTRLVGKALKDLPLPKDAIIGAIVRGNECLIPR 408

Query: 425 GEDVLTVGDRIVVITLLKNITRTSDM 450
           G+  +   DR++V      I +  ++
Sbjct: 409 GDTTIEYEDRLLVFAKWDEIEKIEEI 434
 Identities = 48/212 (22%), Positives = 99/212 (46%), Gaps = 15/212 (7%)

Query: 3   VKMRIIVVGGGKVGTALCRSLVAEKHDVVLIEKKENVLKRVTKQHDIMGIV-GNGANYKI 61
           V  R+ + G G +G+   + L      V LIE      + ++ + + + IV G+  + +
Sbjct: 221 VTRRVTIFGAGTIGSYTAKILAKGMTSVKLIESSMERCEALSGELEGVRIVCGDATDIEF 280

Query: 62  LEQAEVKNCDIFIAITDRDEVNMISAVLAKKMGAKETVVRMRNPEYSNPYFKDKNFLGFS 121
           L + E+    D  +A T+ DE N++ ++L+K +GA+  + ++   EY   +      +G
Sbjct: 281 LIEEEIGKSDAVLAATESDEKNLLISLLSKNLGARIAIAKVEKREYVKLF----EAVGVD 336

Query: 122 SVVNPELLAAQYIA---NTIEFPNATSVEHFANGRVMLMEFKILEGNKLCHTSMSQIRKK 178
             +NP  +    ++     T+        +E  A   V++   +++ G  L    + +
Sbjct: 337 VALNPRSVTYNEVSKLLRTMRIETLAEIEGTAVVEVVVRNTRLV-GKALKDLPLPK---- 391

Query: 179 FGNIVICAIERDGKLIIPDGDATIQVKDKIFV 210
             + +I AI R  + +IP GD TI+ +D++ V
Sbjct: 392 --DAIIGAIVRGNECLIPRGDTTIEYEDRLLV 421
```

There is also homology to SEQ ID 4716.

SEQ ID 8836 (GBS384) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 2; MW 53kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 6; MW 78kDa).

The GBS384-GST fusion product was purified (Figure 212, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 279), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1532

A DNA sequence (GBSx1623) was identified in *S.agalactiae* <SEQ ID 4717> which encodes the amino acid sequence <SEQ ID 4718>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4948(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1533

A DNA sequence (GBSx1624) was identified in *S.agalactiae* <SEQ ID 4719> which encodes the amino acid sequence <SEQ ID 4720>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.58    Transmembrane   37 -  53 (  33 -  61)
   INTEGRAL    Likelihood =-11.57    Transmembrane  183 - 199 ( 179 - 214)
   INTEGRAL    Likelihood =-10.03    Transmembrane  397 - 413 ( 392 - 424)
   INTEGRAL    Likelihood = -6.79    Transmembrane   14 -  30 (   5 -  31)
   INTEGRAL    Likelihood = -6.42    Transmembrane   71 -  87 (  69 -  93)
   INTEGRAL    Likelihood = -4.99    Transmembrane  278 - 294 ( 274 - 295)
   INTEGRAL    Likelihood = -4.19    Transmembrane  133 - 149 ( 132 - 152)
   INTEGRAL    Likelihood = -4.09    Transmembrane  327 - 343 ( 324 - 344)
   INTEGRAL    Likelihood = -2.44    Transmembrane  236 - 252 ( 234 - 252)
   INTEGRAL    Likelihood = -0.59    Transmembrane  456 - 472 ( 456 - 472)

----- Final Results -----
               bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10065> which encodes amino acid sequence <SEQ ID 10066> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB90400 GB:AE001046 TRK potassium uptake system protein (trkH)
           [Archaeoglobus fulgidus]
 Identities = 166/480 (34%), Positives = 262/480 (54%), Gaps = 10/480 (2%)

Query: 1   MNKSMIRFLLSKLLLIEAALLAIPLTVGLIYREP-QSVMMSIVITMIILIILGLLGSLFK 59
           MN  +   +L KLL++ +    +PL    ++ EP    ++   +++++  +LG  G   +
Sbjct: 1   MNLRLTASILGKLLMLFSFSFILPLIAAHVFEEPYHPFLIPAALSLLVGAVLGY-GIKTE 59

Query: 60  PKNYHIYTKEGMLIVALCWILWSFFGALPFVISGQIPNIIDAFFEVSSGFTTTGATILDD 119
            +   +  KE   IVAL W+  S FG++P++I G  P  +DAFFE  SGFTTTGA++L
Sbjct: 60  SEFDSLRHKESFAIVALIWLFMSIFGSIPYIIFGISP--VDAFFESMSGFTTTGASVLTP 117

Query: 120 VSVLSPALLFWRSFTHLIGGMGVLVFALAIMENSKNSHLEVMRAEVPGPVFGKVVSKLKK 179
              L  +LL WRS T  IGGMG++V  LAI  N       + +AE PG    K+  +++
Sbjct: 118 EE-LPKSLLLWRSLTQWIGGMGIIVLFLAIFPNVAKRSTVLFQAEYPGVSLSKLKPRIRD 176

Query: 180 TAQILYLLYLLMFAVFAVILYFAGMPFFDSIIIAMGTAGTGGFAVYNDSIAHYNSPLITN 239
           TA  LY +YLL+      +LY  G+  FD+I     T  TGG++ +++SIA +    +
Sbjct: 177 TALSLYKVYLLLTIAEVALLYALGLSLFDAINHTFTTLSTGGYSTHSESIAFFKDVRVEA 236

Query: 240 LVSIGMLIFGVNFNLYYLLLLRKIKAFFGDEELKTYLRIVAIATFMIALNVIGMYDNFRQ 299
           +V+    + G NF L Y LL  K    F + E + Y+  +A+A+ +IA   +  Y  F +
Sbjct: 237 VVAFFAFLGGANFALIYFLLSGK-PVIFRNTEFRAYVCFLALASVVIAAVNLDRYSIF-E 294

Query: 300 GLEHIFFEVSAIITTTGFGVTDITRWPLFSQVILLFLMFIGGSAGSTAGGFKVMRSLILA 359
            L +  F+  +I+TTTGF   D   W   +++IL+ LMFIGGS+GST GG KV+R  +L
Sbjct: 295 SLRYSIFQAVSIMTTTGFTTADFDAWSDSAKLILVVLMFIGGSSGSTGGGIKVIRIYLLI 354

Query: 360 KIARNQVLSTLYPNRVMSLHINKSVLDKNTQHGVLKYLTIYLAIFMALVLVLTLDTNDFL 419
           K A +Q+L    P  V ++      + K     +  +  +Y+ IF    ++++L   D +
Sbjct: 355 KYAVHQILRAAEPRTVRAVKFEGRAIKKEILDDIAAFFVLYILIFAVSSILVSLSGYDIV 414

Query: 420 VVISAAASCFNNIGP---LLGSNETFSFFSPFSKLLLSFAMIAGRLEIYPVLLMFIPKTW 476
             ISA A+   N+GP   L G+ E ++ F   +K+LL+  M  GRLEI+ V+ +FIP  W
Sbjct: 415 TSISATAATLGNVGPGLGLAGAAENYASFPSLTKILLAVNMWIGRLEIFTVVSLFIPTFW 474
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1534

A DNA sequence (GBSx1625) was identified in *S.agalactiae* <SEQ ID 4721> which encodes the amino acid sequence <SEQ ID 4722>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence (or aa 1-20)

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2870(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD36530 GB:AE001797 conserved hypothetical protein
         [Thermotoga maritima]
 Identities = 43/75 (57%), Positives = 57/75 (75%), Gaps = 1/75 (1%)

Query: 2  LKSFLIFLVRFYQKNISPAFPASCRYRPTCSTYMIEAIQKHG-LKGVLMGIARILRCHPL 60
          +K  LI L+RFYQ+ ISP  P +CR+ PTCS Y I+A++KHG LKG  +G+ RILRC+PL
Sbjct: 1  MKKLLIMLIRFYQRYISPLKPPTCRFTPTCSNYFIQALEKHGLLKGTFLGLRRILRCNPL 60

Query: 61 AHGGNDPVPDHFSLR 75
          + GG DPVP+ FS +
Sbjct: 61 SKGGYDPVPEEFSFK 75
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4723> which encodes the amino acid sequence <SEQ ID 4724>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3639(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 53/78 (67%), Positives = 60/78 (75%)

Query: 1  MLKSFLIFLVRFYQKNISPAFPASCRYRPTCSTYMIEAIQKHGLKGVLMGIARILRCHPL 60
          M+K  LI  V+ YQK ISP  P SCRY+PTCS YM+ AI+KHG KG+LMGIARILRCHP
Sbjct: 1  MMKKLLIVSVKAYQKYISPLSPPSCRYKPTCSAYMLTAIEKHGTKGILMGIARILRCHPF 60

Query: 61 AHGGNDPVPDHFSLRRNK 78
            GG DPVP+ FSL RNK
Sbjct: 61 VAGGVDPVPEDFSLMRNK 78
```

SEQ ID 4722 (GBS233) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 3; MW 35.6kDa).

The GBS233-GST fusion product was purified (Figure 207, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 280), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1535

A DNA sequence (GBSx1626) was identified in *S.agalactiae* <SEQ ID 4725> which encodes the amino acid sequence <SEQ ID 4726>. This protein is predicted to be ribosomal large subunit pseudouridine synthase B (rluB). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2957(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05295 GB:AP001512 pseudouridylate synthase [Bacillus halodurans]
 Identities = 130/239 (54%), Positives = 175/239 (72%), Gaps = 2/239 (0%)

Query: 2   RINKYIAHAGIASRRKAEELIKQGMVTINGQVVNELATQVKAG-DLVEIEGSPIYNEEKV 60
           R+ K IA AGIASRRKAE+LI +G V +NGQVV EL  +V    D +E+EG P+  EE V
Sbjct: 3   RLQKVIAQAGIASRRKAEQLILEGKVKVNGQVVKELGIKVNPNQDDIEVEGVPVEKEEPV 62

Query: 61  YYLLNKPRGVISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTTGLLILTNDGDFTDKM 120
           Y+LL KP GVISSV DDKGRK V D L ++++R+YPVGRLD+DT+GLL+LTNDG+F + +
Sbjct: 63  YFLLYKPTGVISSVKDDKGRKVVTDFL-EIEQRVYPVGRLDYDTSGLLLLTNDGEFANLL 121

Query: 121 IHPRNEIDKVYLARVKGIATKENLRPLTRGVVIDGKKTKPARYTIIKVDHEKNRSVVELT 180
           +HPR++I+KVY+A+VKGI T++ L+ L RGV ++    T PA+  ++ VD  K  ++V+LT
Sbjct: 122 MHPRHKIEKVYVAKVKGIPTRDQLKLLARGVKLEDGPTAPAKVKMLSVDRRKQTAIVKLT 181

Query: 181 IHEGRNHQVKKMFEQVGLLVDKLSRTQFGTLDLTGLRPGEARRLNKKEISQLHNAAINK 239
           IHEGRN QV++MFE +G  V KL R QF  LDL+G+ PG+ R L   E+  L   A+ K
Sbjct: 182 IHEGRNRQVRRMFETIGCEVMKLKREQFAFLDLSGMNPGDVRPLKPIEVKHLRELAVTK 240
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4727> which encodes the amino acid sequence <SEQ ID 4728>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1587(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 210/239 (87%), Positives = 228/239 (94%)

Query: 1   MRINKYIAHAGIASRRKAEELIKQGMVTINGQVVNELATQVKAGDLVEIEGSPIYNEEKV 60
           MRINKYIAHAGIASRRKAEELIKQG+VT+NGQV+ +LAT VK+GD+VEIEGSPIYNEEKV
Sbjct: 9   MRINKYIAHAGIASRRKAEELIKQGLVTLNGQVITDLATTVKSGDVVEIEGSPIYNEEKV 68

Query: 61  YYLLNKPRGVISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTTGLLILTNDGDFTDKM 120
           YYLLNKPRG ISSVSDDKGRKTV+DLLPQVKERIYPVGRLDWDT+G+LILTNDGDFTD M
Sbjct: 69  YYLLNKPRGAISSVSDDKGRKTVLDLLPQVKERIYPVGRLDWDTSGVLILTNDGDFTDTM 128

Query: 121 IHPRNEIDKVYLARVKGIATKENLRPLTRGVVIDGKKTKPARYTIIKVDHEKNRSVVELT 180
           IHPRNEIDKVYLARVKGIATKENLRPLTRG+VIDGKKTKPARY I++V+ +K+RS+VELT
Sbjct: 129 IHPRNEIDKVYLARVKGIATKENLRPLTRGIVIDGKKTKPARYNIVRVEADKSRSIVELT 188

Query: 181 IHEGRNHQVKKMFEQVGLLVDKLSRTQFGTLDLTGLRPGEARRLNKKEISQLHNAAINK 239
           IHEGRNHQVKKMFE VGLLVDKLSRT+FGT+DL GLRPGEARRLNKKEISQLHN A  K
Sbjct: 189 IHEGRNHQVKKMFESVGLLVDKLSRTRFGTVDLKGLRPGEARRLNKKEISQLHNLANTK 247
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1536

A DNA sequence (GBSx1627) was identified in *S.agalactiae* <SEQ ID 4729> which encodes the amino acid sequence <SEQ ID 4730>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1476(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05280 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 75/180 (41%), Positives = 107/180 (58%), Gaps = 10/180 (5%)

Query: 6   SIEALLFVAGEDGLSLRQMAELLSLTPSALIQQLEKLAKRYEEDDDSSLLLLETAQTYKL 65
           +IE +LFV G++G++L ++ +LL L+   +   LE+L   Y  D+   L + E A  ++L
Sbjct: 9   AIEGILFVRGDEGVTLEELCDLLELSTDVVYAALEELRLSYT-DEARGLRIEEVAHAFRL 67

Query: 66  VTKDSYMTLLRDYAKAPINQSLSRASLEVLSIIAYKQPITRIEIDDIRGVNSSGAITRLI 125
            TK       +  A + +   LS+A+LE L+IIAY+QPITRIE+D++RGV S  AI  L
Sbjct: 68  STKPELAPYFKKLALSTLQSGLSQAALETLAIIAYRQPITRIEVDEVRGVKSEKAIQTLT 127

Query: 126 AFGLIKEAGKKEVLGRPNLYETTNYFLDYMGINQLDDL------IDASSIELVDEEVSLF 179
           +  LIKE G+ +  GRP LY TT  FLD+ G+  L +L      ID SSI    EE  LF
Sbjct: 128 SRLLIKEVGRAQGTGRPILYGTTPQFLDHFGLKSLKELPPLPEDIDESSI---GEEADLF 184
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4731> which encodes the amino acid sequence <SEQ ID 4732>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1062(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 130/179 (72%), Positives = 159/179 (88%)

Query: 1   MTYLGSIEALLFVAGEDGLSLRQMAELLSLTPSALIQQLEKLAKRYEEDDDSSLLLLETA 60
           MTYL  IEALLFVAGE+GLSLR +A +LSLTP+AL QQLEKL+++YE+D  SSL L+ETA
Sbjct: 1   MTYLSQIEALLFVAGEEGLSLRHLASMLSLTPTALQQQLEKLSQKYEKDQHSSLCLIETA 60

Query: 61  QTYKLVTKDSYMTLLRDYAKAPINQSLSRASLEVLSIIAYKQPITRIEIDDIRGVNSSGA 120
            TY+LVTK+ +  LLR YAK P+NQSLSRASLEVLSI+AYKQPITRIEIDDIRGVNSSGA
Sbjct: 61  NTYRLVTKEGFAELLRAYAKTPMNQSLSRASLEVLSIVAYKQPITRIEIDDIRGVNSSGA 120

Query: 121 ITRLIAFGLIKEAGKKEVLGRPNLYETTNYFLDYMGINQLDDLIDASSIELVDEEVSLF 179
           +++L+AF LI+EAGKK+V+GRP+LY TT+YFLDYMGIN LD+LI+ S++E  DEE++LF
Sbjct: 121 LSKLLAFDLIREAGKKDVVGRPHLYATTDYFLDYMGINHLDELIEVSAVEPADEEIALF 179
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1537

A DNA sequence (GBSx1628) was identified in *S.agalactiae* <SEQ ID 4733> which encodes the amino acid sequence <SEQ ID 4734>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1012(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 89/246 (36%), Positives = 145/246 (58%), Gaps = 19/246 (7%)

Query: 3   IKLKDFEGPLDLLLHLVSKYEVDIYDVPIVEVIEQYLAYIATLQAMRLEVAGEYMLMASQ 62
           +K+  FEGPLDLLLHL+++ E+DIYD+P+ ++ EQYL Y+ T++ + L++A EY++MA+
Sbjct: 6   VKIDTFEGPLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65

Query: 63  LMLIKSRNLLPK----VVESNPI-EDDPEMELLSQLEEYRRFKVLSEELANQHQERAKYF 117
           L+ IKSR LLPK    + E   + E+DP  EL+ +L EYR++K  +++L  + +ER K F
Sbjct: 66  LLSIKSRMLLPKQEEELFEDELLEEEDPREELIEKLIEYRKYKDAAKDLKEREEERQKSF 125

Query: 118 SKPKQEVIFEDAILLHDKSVMDLFLTFSQMMSQKQKELSNS------QTVIEKEDYRIED 171
           +KP  ++   +      +S   L +T   M+   QK L         +T I ++D  IE
Sbjct: 126 TKPPSDL--SEYAKEVKQSEQKLSVTVYDMIGAFQKVLKRKKINRPMETTITRQDIPIEA 183

Query: 172 MMIVIERHFNLKKKTT---LQEVFADCQTKSEMITLFLAMLELIKLHQITVEQDSNFSQV 228
            M  I    +LK + T     ++F   + K  ++  FLA+LEL+K   + +EQ+ NFS +
Sbjct: 184 RMNEIVH--SLKSRGTRINFMDLF-PYEQKEHLVVTFLAVLELMKNQLVLIEQEHNFSDI 240

Query: 229 ILRKEE 234
            +   E
Sbjct: 241 YITGSE 246
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4735> which encodes the amino acid sequence <SEQ ID 4736>. Analysis of this protein sequence reveals the following:

```
     Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.61    Transmembrane  199 - 215 ( 199 - 218)

----- Final Results -----
               bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 86/239 (35%), Positives = 145/239 (59%), Gaps = 15/239 (6%)

Query: 3   IKLKDFEGPLDLLLHLVSQYKVDIYEVPIVEVIEQYLNYIETLQVMKLEVAGDYMLMASQ 62
           +K+  FEGPLDLLLHL+++ ++DIY++P+ ++ EQYL Y+ T++V++L++A +Y++MA+
Sbjct: 6   VKIDTFEGPLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65

Query: 63  LMLIKSRRLLPKVVEHI-------EEDLEQDLLEKIEEYSRFKAVSQALAKQHDQRAKWY 115
           L+ IKSR LLPK  E +       EED  ++L+EK+ EY ++K  ++ L ++ ++R K +
Sbjct: 66  LLSIKSRMLLPKQEEELFEDELLEEEDPREELIEKLIEYRKYKDAAKDLKEREEERQKSF 125

Query: 116 SKPKQELI-FEDAILQEDK----TVMDLFLAFSNIMAAKRAVLKNNHTVIERDDYKIEDM 170
           +KP  +L  +   + Q ++    TV D+  AF  ++  K+ + +   T I R D  IE
Sbjct: 126 TKPPSDLSEYAKEVKQSEQKLSVTVYDMIGAFQKVLKRKK-INRPMETTITRQDIPIEAR 184
```

```
Query: 171 MASIKQRLEKENV-IRLSAIFEECQTLNEVISIFLASLELIKLHVVFVEQLSNFGAIIL 228
           M  I   L+     I    +F   Q  + V++ FLA LEL+K  +V +EQ  NF  I +
Sbjct: 185 MNEIVHSLKSRGTRINFMDLFPYEQKEHLVVT-FLAVLELMKNQLVLIEQEHNFSDIYI 242
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 156/235 (66%), Positives = 191/235 (80%), Gaps = 2/235 (0%)

Query: 1   MDIKLKDFEGPLDLLLHLVSKYEVDIYDVPIVEVIEQYLAYIATLQAMRLEVAGEYMLMA 60
           MDIKLKDFEGPLDLLLHLVS+Y+VDIY+VPIVEVIEQYL YI TLQ M+LEVAG+YMLMA
Sbjct: 1   MDIKLKDFEGPLDLLLHLVSQYKVDIYEVPIVEVIEQYLNYIETLQVMKLEVAGDYMLMA 60

Query: 61  SQLMLIKSRNLLPKVVESNPIEDDPEMELLSQLEEYRRFKVLSEELANQHQERAKYFSKP 120
           SQLMLIKSR LLPKVVE   IE+D E +LL ++EEY RFK +S+ LA QH +RAK++SKP
Sbjct: 61  SQLMLIKSRRLLPKVVEH--IEEDLEQDLLEKIEEYSRFKAVSQALAKQHDQRAKWYSKP 118

Query: 121 KQEVIFEDAILLHDKSVMDLFLTFSQMMSQKQKELSNSQTVIEKEDYRIEDMMIVIERHF 180
           KQE+IFEDAIL  DK+VMDLFL FS +M+ K+  L N+ TVIE++DY+IEDMM  I++
Sbjct: 119 KQELIFEDAILQEDKTVMDLFLAFSNIMAAKRAVLKNNHTVIERDDYKIEDMMASIKQRL 178

Query: 181 NLKKKTTLQEVFADCQTKSEMITLFLAMLELIKLHQITVEQDSNFSQVILRKEEK 235
             +    L  +F +CQT +E+I++FLA LELIKLH + VEQ SNF  +ILRKE+K
Sbjct: 179 EKENVIRLSAIFEECQTLNEVISIFLASLELIKLHVVFVEQLSNFGAIILRKEKK 233
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1538

A DNA sequence (GBSx1629) was identified in *S.agalactiae* <SEQ ID 4737> which encodes the amino acid sequence <SEQ ID 4738>. This protein is predicted to be pXO1-18. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.14   Transmembrane  128 - 144 ( 127 - 145)

----- Final Results -----
               bacterial membrane --- Certainty=0.2657(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05248 GB:AP001512 integrase/recombinase [Bacillus halodurans]
 Identities = 67/271 (24%), Positives = 117/271 (42%), Gaps = 35/271 (12%)

Query: 11  LKTMINDINNFIESKK----LSLNSRKSYHYDLKQFYKII--------GGHVNSEKLALY 58
           ++T+ N++  F+  +K    LS N+ +SY  DLKQ+ + +          ++  E +  Y
Sbjct: 1   METVNNNLQQFLHFQKVERGLSNNTIQSYGRDLKQYIQYVERVEEIRSARNITRETILHY 60

Query: 59  QQSLSEFKL--TARKRKLSAVNQFLFFLYNRGTLKEFYRL-----QETEKITLAQTKSQI 111
              L E     T+  R ++A+  F  FL           +     + T+++  A T  ++
Sbjct: 61  LYHLREQGRAETSIARAVAAIRSFHQFLLREKLSDSDPTVHVEIPKATKRLPKALTIEEV 120

Query: 112 MDLSNFYQDTDYPSGRLIALLIL--SLGLTPAEIANLKKADFDTTFNILS-IEKSQMKRI 168
             L N  Q  D  S R  A+L L  + G+  +E+  L  +D   +   +  + K   +RI
Sbjct: 121 EALLNSPQGRDPFSLRNKAMLELLYATGMRVSELIGLTLSDIHLSMGFVRCLGKGNKERI 180

Query: 169 LKLPEDLLPFLLESLEEDG----------DLVF-EHNGKPYSRQWFFNQLTDFLNEKN-E 216
           + + + +    +ES   +G          D VF  H+G+P SRQ F+  L       N +
Sbjct: 181 IPIGQ-VATEAVESYLANGRGKLMKKQSHDHVFVNHHGRPLSRQGFWKMLKQLAKNVNID 239

Query: 217 QQLTAQLLREQFILKQKENGKTMTELSRLLG 247
```

```
           + LT   LR  F     ENG  +  +  +LG
Sbjct: 240 KPLTPHTLRHSFATHLLENGADLRAVQEMLG 270
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4739> which encodes the amino acid sequence <SEQ ID 4740>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.90   Transmembrane  111 - 127 ( 110 - 127)

----- Final Results -----
               bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 117/243 (48%), Positives = 167/243 (68%), Gaps = 1/243 (0%)

Query: 18  INNFIESKKLSLNSRKSYHYDLKQFYKIIGGHVNSEKLALYQQSLSEFKLTARKRKLSAV 77
            I  FI SK LS NS+K+Y YDL+QF ++IG  VN +KL LYQ S++   L+A+KRKLS
Sbjct: 5   IEPFIASKALSQNSQKAYRYDLQQFCQLIGERVNQDKLLLYQNSIANLSLSAKKRKLSTA 64

Query: 78  NQFLFFLYNRGTLKEFYRLQETEKITLAQTK-SQIMDLSNFYQDTDYPSGRLIALLILSL 136
            NQFL++LY    L  ++RL +T K+   + + + I++   FYQ T +  G+LI+LLIL L
Sbjct: 65  NQFLYYLYQIKYLNSYFRLTDTMKVMRTEKQQAAIINTDIFYQKTPFVWGQLISLLILEL 124

Query: 137 GLTPAEIANLKKADFDTTFNILSIEKSQMKRILKLPEDLLPFLLESLEEDGDLVFEHNGK 196
            GLTP+E+A ++ A+ D  F +L+++  +  R+L L + L+PFL + L      +FEH G
Sbjct: 125 GLTPSEVAGIEVANLDLNFQMLTLKTKKGVRVLPLSQILIPFLEQQLVGKEVYLFEHRGI 184

Query: 197 PYSRQWFFNQLTDFLNEKNEQQLTAQLLREQFILKQKENGKTMTELSRLLGLKTPITLER 256
            P+SRQWFFN L  F+      + LTAQ LREQFILK+K  GK++ ELS +LGLK+P+TLE+
Sbjct: 185 PFSRQWFFNHLKTFVRSIGYEGLTAQKLREQFILKEKLAGKSIIELSDILGLKSPMTLEK 244

Query: 257 YYR 259
            YY+
Sbjct: 245 YYK 247
```

SEQ ID 4738 (GBS383) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 7; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 5; MW 57.1kDa).

The GBS383-GST fusion product was purified (Figure 212, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 308), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1539

A DNA sequence (GBSx1630) was identified in *S.agalactiae* <SEQ ID 4741> which encodes the amino acid sequence <SEQ ID 4742>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2465(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05201 GB:AP001512 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 38/136 (27%), Positives = 73/136 (52%), Gaps = 1/136 (0%)

Query: 7   ESFLLNHLDHYLIPAEDVAIFVDTHNADHVMLLLASNGFSRVPVITKEKKYVGTISISDI 66
           ++ + N L   +IP E VA    ++  +H +L+L  +G++ +PV+ +  K  G IS S I
Sbjct: 7   QNIMDNDLKELVIPFEKVAHVHLSNPLEHALLVLIKSGYTAIPVLDEHSKLHGVISKSLI 66

Query: 67  MGYQSKGQLTDWE-MAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISDQNDFR 125
           +      +  + E +A   + +++N +I  I+  A+ +  +   + +PF+ ++ D   F
Sbjct: 67  LDALLGVERIEMERLAHLVVKDVMNPEIPTIHHKASFSRALKVSIAHPFICILDDDGSFL 126

Query: 126 GIITRKSILKAINSLL 141
           GI+TR +IL  IN  L
Sbjct: 127 GILTRSTILSFINRQL 142
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4743> which encodes the amino acid sequence <SEQ ID 4744>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3539(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 119/153 (77%), Positives = 137/153 (88%)

Query: 1   MIAKEFESFLLNHLDHYLIPAEDVAIFVDTHNADHVMLLLASNGFSRVPVITKEKKYVGT 60
           MIAKEFE+FL++HLD+YLIP +D+AIF+DTHNADHVMLLL SNGFSRVPVIT+EKKYVGT
Sbjct: 1   MIAKEFETFLMSHLDNYLIPEQDLAIFIDTHNADHVMLLLVSNGFSRVPVITREKKYVGT 60

Query: 61  ISISDIMGYQSKGQLTDWEMAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISD 120
           ISISDIM YQSK QLTDWEM+QTDI EMVNTKIE I+  ++LT IMHK++D+PFLPV+
Sbjct: 61  ISISDIMMYQSKRQLTDWEMSQTDIGEMVNTKIETISITSSLTEIMHKLIDFPFLPVVDR 120

Query: 121 QNDFRGIITRKSILKAINSLLHDFTDEYTITPK 153
            N F GIITRKSILKA+NSLLHDFTD+YTI  K
Sbjct: 121 ANRFVGIITRKSILKAVNSLLHDFTDDYTIIKK 153
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1540

A DNA sequence (GBSx1631) was identified in *S.agalactiae* <SEQ ID 4745> which encodes the amino acid sequence <SEQ ID 4746>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4421(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06785 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 55/169 (32%), Positives = 95/169 (55%), Gaps = 1/169 (0%)

Query: 5   KLVVMSDSHGDRDIVKDIKNHYLGKVDAIFHNGDSELPSSDPIWEGIHVVTGNCDYDSGY 64
           KL+++SDSHG  D +K + + +  +VDAI H GDSELP  D   EG+++V GNCD+   +
Sbjct: 2   KLLILSDSHGWSDELKAVADKHRQEVDAIIHCGDSELPRDDRALEGMNIVRGNCDFGVDF 61

Query: 65  PEVLVTKIDNAVIVQTHGHLHQINFTWDKLDLLAQQEDADICLYGHLHRADAWKNGKTIF 124
           PE  +  + +  +  THGHL+ +  ++  L   A++  A +  +GH H A +++    +F
Sbjct: 62  PEDFIKTVGDFNVYVTHGHLYNVKMSYVSLTYRAEEVGAQLVCFGHSHVATSFQENGIVF 121

Query: 125 INPGSVLQPRGPINEKLYAVVTITDSKVLVEYYTRQHQPYPNLTKELSR 173
           +NPGS+  PR    E+ Y +  + D ++ + +  R      +L +   R
Sbjct: 122 VNPGSLRLPRNR-KEQTYCLAYVRDDQIELTFLDRDGHEVTDLQRTYLR 169
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4747> which encodes the amino acid sequence <SEQ ID 4748>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3835(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 117/173 (67%), Positives = 143/173 (82%)

Query: 1   MAIRKLVVMSDSHGDRDIVKDIKNHYLGKVDAIFHNGDSELPSSDPIWEGIHVVTGNCDY 60
           MA + ++VMSDSHGDRDIV+ IK+ YLG+VDAIFHNGDSEL SSDPIW GI+VV GNCDY
Sbjct: 1   MASKTIIVMSDSHGDRDIVQAIKDKYLGQVDAIFHNGDSELNSSDPIWAGIYVVGGNCDY 60

Query: 61  DSGYPEVLVTKIDNAVIVQTHGHLHQINFTWDKLDLLAQQEDADICLYGHLHRADAWKNG 120
           D+GYP+ LVT++    I QTHGHL+ INFTWDKLD  AQ+  ADICLYGHLHR  AW+ G
Sbjct: 61  DTGYPDRLVTQLGTVTIAQTHGHLYHINFTWDKLDYFAQEVVADICLYGHLHRPAAWQVG 120

Query: 121 KTIFINPGSVLQPRGPINEKLYAVVTITDSKVLVEYYTRQHQPYPNLTKELSR 173
           +T+F+NPGSV QPRG INEKLYA V +TD+++ V+Y+TR H+ YP+L+KE  R
Sbjct: 121 QTLFMNPGSVTQPRGEINEKLYARVELTDTQIKVDYFTRDHKLYPSLSKEFKR 173
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1541

A DNA sequence (GBSx1632) was identified in *S.agalactiae* <SEQ ID 4749> which encodes the amino acid sequence <SEQ ID 4750>. This protein is predicted to be HAM1 family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1218(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14796 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
```

```
 Identities = 96/189 (50%), Positives = 130/189 (67%), Gaps = 1/189 (0%)

Query: 128 LIATHNEGKTKEFRELFGKLGLKVENLNDYPDLPEVEETGMTFEENARLKAETISKLTGK 187
           +IATHN GK KEF+E+    G  V++L +     E+EETG TFEENA +KAE ++K   K
Sbjct: 8   IIATHNPGKVKEFKEILEPRGYDVKSLAEIGFTEEIEETGHTFEENAIMKAEAVAKAVNK 67

Query: 188 MVISDDSGLKVDALGGLPGVWSARFSGPDATDARNNAKLLHELAMVFDKERRSAQFHTTL 247
           MVI+DDSGL +D LGG PGV+SAR++G    D  N  K+L EL  + +KE+R+A+F   L
Sbjct: 68  MVIADDSGLSIDNLGGRPGVYSARYAGEQKDDQANIEKVLSELKGI-EKEQRTARFRCAL 126

Query: 248 VVSAPNKESLVVEAEWPGYIGTEPKGENGFGYDPLFIVGEGSRTAAELSAQEKNNLSHRG 307
            VS P +E+  VE    GYI  EP+GE GFGYDP+FIV +  +T AEL++ EKN +SHR
Sbjct: 127 AVSIPGEETKTVEGHVEGYIAEEPRGEYGFGYDPIFIVKDKDKTMAELTSDEKNKISHRA 186

Query: 308 QAVRKLMEV 316
            A++KL ++
Sbjct: 187 DALKKLSKL 195
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4751> which encodes the amino acid sequence <SEQ ID 4752>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 214/325 (65%), Positives = 253/325 (77%), Gaps = 5/325 (1%)

Query: 1   MTKTIFESKTEGNWFLGSFQAFNYFTCFG-NDESYEAIQDVFHRLLSTLKVE---GLQLH 56
           M++ I+E K E NWF+G     N  + +G    + + I D+   + +TL  E   G  +
Sbjct: 14  MSEKIYEYKDENNWFIGKMTGHNLISGWGVKHTTIKKIDDLLDGIAATLDWENPKGYDVS 73

Query: 57  VVQMTSDFQLLAFLVDMINQEYSRHIKVTQHKGAILVSEDDQLFLVHLPKEGTSLEKFFD 116
           VV+  S   L+ F++DMINQE  R IKVT H G IL+ E+ +L  V+LP+ G S   FF
Sbjct: 74  VVRHQSPLSLITFIIDMINQETQREIKVTPHAGTILLMENAKLLAVYLPEGGVSTATFF- 132

Query: 117 LKNDNNFGDTILIATHNEGKTKEFRELFGKLGLKVENLNDYPDLPEVEETGMTFEENARL 176
             ++  FGD ILIAT NEGKTKEFR LFG+LG +VENLNDYP+LPEV ETG TFEENARL
Sbjct: 133 ATSEQGFGDIILIATRNEGKTKEFRNLFGQLGYRVENLNDYPELPEVAETGTTFEENARL 192

Query: 177 KAETISKLTGKMVISDDSGLKVDALGGLPGVWSARFSGPDATDARNNAKLLHELAMVFDK 236
           KAETIS+LTGKMV++DDSGLKVDALGGLPGVWSARFSGPDATDA+NNAKLLHELAMVFD+
Sbjct: 193 KAETISRLTGKMVLADDSGLKVDALGGLPGVWSARFSGPDATDAKNNAKLLHELAMVFDQ 252

Query: 237 ERRSAQFHTTLVVSAPNKESLVVEAEWPGYIGTEPKGENGFGYDPLFIVGEGSRTAAELS 296
           ++RSAQFHTTLVV+APNK+SLVVEA+WPGYI T+PKGENGFGYDP+FIVGE    AAEL
Sbjct: 253 KKRSAQFHTTLVVAAPNKDSLVVEADWPGYIATQPKGENGFGYDPVFIVGETGHHAAELE 312

Query: 297 AQEKNNLSHRGQAVRKLMEVFPKWQ 321
           A +KN LSHRGQAVRKLMEVFP WQ
Sbjct: 313 ADQKNQLSHRGQAVRKLMEVFPAWQ 337
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1542

A DNA sequence (GBSx1633) was identified in *S.agalactiae* <SEQ ID 4753> which encodes the amino acid sequence <SEQ ID 4754>. This protein is predicted to be glutamate racemase (murI). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.86   Transmembrane  114 - 130 ( 114 - 130)

----- Final Results -----
              bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10067> which encodes amino acid sequence <SEQ ID 10068> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
 Identities = 160/267 (59%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

Query: 27  MDSRPIGFLDSGVGGLTVVKEMFRQLPEEEVIFIGDQARAPYGPRPAQQIREFTWQMVNF 86
            M  + IGF+DSGVGGLTVVKE  RQLP E + ++GD AR PYGPRP  Q+R+FTW+M +F
Sbjct: 1   MKKQAIGFIDSGVGGLTVVKEAMRQLPNESIYYVGDTARCPYGPRPEDQVRKFTWEMTHF 60

Query: 87  LLTKNVKMIVIACNTATAVAWQEIKEKLDIPVLGVILPGASAAIKSTNLGKVGIIGTPMT 146
            LL KN+KM+VIACNTATA A ++IK+KL IPV+GVILPG+ AAIK+T+  ++G+IGT  T
Sbjct: 61  LLDKNIKMLVIACNTATAAALKDIKKKLAIPVIGVILPGSRAAIKATHTNRIGVIGTEGT 120

Query: 147 VKSDAYRQKIQALSPNTAVVSLACPKFVPIVESNQMSSSLAKKVVYETLSPLVGK-LDTL 205
            VKS+ Y++ I +      V SLACPKFVP+VESN+ SS++AKKVV ETL PL  + LDTL
Sbjct: 121 VKSNQYKKMIHSKDTKALVTSLACPKFVPLVESNEYSSAIAKKVVAETLRPLKNEGLDTL 180

Query: 206 ILGCTHYPLLRPIIQNVMGAEVKLIDSGAETVRDISVLLNYFEINHNWQNKH-GGHHFYT 264
            ILGCTHYPLLRPIIQN +G  V LIDSGAETV ++S +L+YF +   + QNK     +FYT
Sbjct: 181 ILGCTHYPLLRPIIQNTLGDSVTLIDSGAETVSEVSTILDYFNLAVDSQNKEKAERNFYT 240

Query: 265 TASPKGFKEIAEQWLS-QEINVERIVL 290
            T S + F  IA +WL   ++ VE I L
Sbjct: 241 TGSSQMFHAIASEWLQLDDLAVEHITL 267
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4755> which encodes the amino acid sequence <SEQ ID 4756>. Analysis of this protein sequence reveals the following:

```
      Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.70   Transmembrane   88 - 104 (  86 - 104)

----- Final Results -----
              bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
 Identities = 149/267 (55%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

Query: 1   MDTRPIGFLDSGVGGLTVVCELIRQLPHEKIVYIGDSARAPYGPRPKKQIKEYTWELVNF 60
            M  + IGF+DSGVGGLTVV E +RQLP+E I Y+GD+AR PYGPRP+ Q++++TWE+ +F
Sbjct: 1   MKKQAIGFIDSGVGGLTVVKEAMRQLPNESIYYVGDTARCPYGPRPEDQVRKFTWEMTHF 60

Query: 61  LLTQNVKMIVFACNTATAVAWEEVKAALDIPVLGVVLPGASAAIKSTTKGQVGVIGTPMT 120
```

```
           LL +N+KM+V ACNTATA A +++K  L IPV+GV+LPG+ AAIK+T   ++GVIGT  T
Sbjct: 61  LLDKNIKMLVIACNTATAAALKDIKKKLAIPVIGVILPGSRAAIKATHTNRIGVIGTEGT 120

Query: 121 VASDIYRKKIQLLAPSIQVRSLACPKFVPIVESNEMCSSIAKKIVYDSLAPLVGK-IDTL 179
           V S+ Y+K I        V SLACPKFVP+VESNE  S+IAKK+V ++L PL  + +DTL
Sbjct: 121 VKSNQYKKMIHSKDTKALVTSLACPKFVPLVESNEYSSAIAKKVVAETLRPLKNEGLDTL 180

Query: 180 VLGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLLNYFDIN-GNYHQKAVEHRFFT 238
           +LGCTHYPLLRPIIQN +G SV LIDSGAE V ++S +L+YF++   + +++  E  F+T
Sbjct: 181 ILGCTHYPLLRPIIQNTLGDSVTLIDSGAETVSEVSTILDYFNLAVDSQNKEKAERNFYT 240

Query: 239 TANPEIFQEIASIWLK-QKINVEHVTL 264
           T + ++F  IAS WL+   + VEH+TL
Sbjct: 241 TGSSQMFHAIASEWLQLDDLAVEHITL 267
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 195/264 (73%), Positives = 231/264 (86%)

Query: 27  MDSRPIGFLDSGVGGLTVVKEMFRQLPEEEVIFIGDQARAPYGPRPAQQIREFTWQMVNF 86
           MD+RPIGFLDSGVGGLTVV E+ RQLP E++++IGD ARAPYGPRP +QI+E+TW++VNF
Sbjct: 1   MDTRPIGFLDSGVGGLTVVCELIRQLPHEKIVYIGDSARAPYGPRPKKQIKEYTWELVNF 60

Query: 87  LLTKNVKMIVIACNTATAVAWQEIKEKLDIPVLGVILPGASAAIKSTNLGKVGIIGTPMT 146
           LLT+NVKMIV ACNTATAVAW+E+K  LDIPVLGV+LPGASAAIKST   G+VG+IGTPMT
Sbjct: 61  LLTQNVKMIVFACNTATAVAWEEVKAALDIPVLGVVLPGASAAIKSTTKGQVGVIGTPMT 120

Query: 147 VKSDAYRQKIQALSPNTAVVSLACPKFVPIVESNQMSSSLAKKVVYETLSPLVGKLDTLI 206
           V SD YR+KIQ L+P+  V SLACPKFVPIVESN+M SS+AKK+VY++L+PLVGK+DTL+
Sbjct: 121 VASDIYRKKIQLLAPSIQVRSLACPKFVPIVESNEMCSSIAKKIVYDSLAPLVGKIDTLV 180

Query: 207 LGCTHYPLLRPIIQNVMGAEVKLIDSGAETVRDISVLLNYFEINHNWQNKHGGHHFYTTA 266
           LGCTHYPLLRPIIQNVMG  VKLIDSGAE VRDISVLLNYF+IN N+  K   H F+TTA
Sbjct: 181 LGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLLNYFDINGNYHQKAVEHRFFTTA 240

Query: 267 SPKGFKEIAEQWLSQEINVERIVL 290
           +P+ F+EIA  WL Q+INVE + L
Sbjct: 241 NPEIFQEIASIWLKQKINVEHVTL 264
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1543

A DNA sequence (GBSx1634) was identified in *S.agalactiae* <SEQ ID 4757> which encodes the amino acid sequence <SEQ ID 4758>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.36    Transmembrane    3 -  19 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.5543(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13675 GB:Z99113 alternate gene name: yoxG [Bacillus subtilis]
 Identities = 26/72 (36%), Positives = 42/72 (58%)

Query: 1   MSITIWILLIIVALFGGLVGGIFIARKQIEKEIGEHPRLTPDAIREMMSQMGQKPSEAKV 60
           M++ + IL+ +VAL  G+  G FIARK +   + ++P +    +R MM QMG KPS+ K+
Sbjct: 1   MTLWVGILVGVVALLIGVALGFFIARKYMMSYLKKNPPINEQMLRMMMMQMGMKPSQKKI 60

Query: 61 QQTYRNIVKHAK 72
```

```
           Q  + +    K
Sbjct: 61 NQMMKAMNNQTK 72
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4759> which encodes the amino acid sequence <SEQ ID 4760>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.72   Transmembrane    7 -  23 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 62/79 (78%), Positives = 69/79 (86%)

Query: 1  MSITIWILLIIVALFGGLVGGIFIARKQIEKEIGEHPRLTPDAIREMMSQMGQKPSEAKV 60
           MS  IWILL+IVAL  G+ GGIFIARKQIEKEIGEHPRLTP+AIREMMSQMGQKPSEAK+
Sbjct: 1  MSTAIWILLLIVALGVGVFGGIFIARKQIEKEIGEHPRLTPEAIREMMSQMGQKPSEAKI 60

Query: 61 QQTYRNIVKHAKTAIKTKK 79
           QQTYRNI+K +K A+   K
Sbjct: 61 QQTYRNIIKQSKAAVSKGK 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1544

A DNA sequence (GBSx1635) was identified in *S.agalactiae* <SEQ ID 4761> which encodes the amino acid sequence <SEQ ID 4762>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.86   Transmembrane   82 -  98 (  79 - 103)

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1545

A DNA sequence (GBSx1636) was identified in *S.agalactiae* <SEQ ID 4763> which encodes the amino acid sequence <SEQ ID 4764>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.83   Transmembrane   56 -  72 (  50 - 105)
   INTEGRAL    Likelihood = -7.27   Transmembrane   27 -  43 (  17 -  48)
   INTEGRAL    Likelihood = -6.26   Transmembrane   76 -  92 (  73 - 105)
   INTEGRAL    Likelihood = -4.83   Transmembrane  119 - 135 ( 118 - 141)
   INTEGRAL    Likelihood = -1.65   Transmembrane  160 - 176 ( 160 - 176)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8837> which encodes amino acid sequence <SEQ ID 8838> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4765> which encodes the amino acid sequence <SEQ ID 4766>. Analysis of this protein sequence reveals the following:

```
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.99   Transmembrane   45 -  61 (  37 -  94)
   INTEGRAL    Likelihood = -7.06   Transmembrane   74 -  90 (  62 -  94)
   INTEGRAL    Likelihood = -3.45   Transmembrane  110 - 126 ( 108 - 130)
   INTEGRAL    Likelihood = -2.18   Transmembrane  149 - 165 ( 149 - 165)
   INTEGRAL    Likelihood = -1.91   Transmembrane   21 -  37 (  20 -  37)

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 111/173 (64%), Positives = 145/173 (83%)

Query: 12  MSKKTTQMVSYTSILVAFAIMIPIIMPAKIIIGPASFTLASHVPLFLSIFISVPVAILVA 71
            M+KK TQ+++YTSILVAFAI+IPIIMP K+IIGPASFTLASHVPLFL+IF+S+PVAILVA
Sbjct: 1   MTKKPTQLIAYTSILVAFAILIPIIMPLKLIIGPASFTLASHVPLFLAIFMSIPVAILVA 60

Query: 72  LGTGLGFLLAGFPIVIVLRALSHIGFALIAAFLIKSKPSLLMSKWQTLLFAVAINIIHGL 131
            LGT LGFLLAG P++IVLRALSH+ FA++AA+ +  KP L+ S  +   FA  IN+IHGL
Sbjct: 61  LGTTLGFLLAGLPLIIVLRALSHLLFAILAAWWLSRKPQLMTSAVKCFSFAFFINVIHGL 120

Query: 132 LEFITVYIITMTSNSSSTYLWSLFSLIGLGSLLHGLVDFYIALFIWKWMTQKL 184
             EF+ VYI+T T+ +S +Y WS+  LIGLGSL+HG++DFY+AL +W+++ + L
Sbjct: 121 AEFLVVYILTATTATSMSYFWSMLGLIGLGSLIHGILDFYLALVLWRFLAKNL 173
```

A related GBS gene <SEQ ID 10789> and protein <SEQ ID 10790> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   24
     Peak Value of UR:   3.16
     Net Charge of CR: 2
McG: Discrim Score:     12.56
GvH: Signal Score (-7.5): -0.16
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 20
ALOM program   count: 5 value: -10.83 threshold:  0.0
   INTEGRAL    Likelihood =-10.83   Transmembrane   45 -  61 (  39 -  94)
   INTEGRAL    Likelihood = -6.26   Transmembrane   65 -  81 (  62 -  94)
   INTEGRAL    Likelihood = -4.83   Transmembrane  108 - 124 ( 107 - 130)
   INTEGRAL    Likelihood = -1.65   Transmembrane  149 - 165 ( 149 - 165)
   INTEGRAL    Likelihood = -0.27   Transmembrane   24 -  40 (  24 -  40)
   PERIPHERAL  Likelihood =  0.42      86
```

```
 modified ALOM score:   2.67
icml HYPID: 7  CFP: 0.533

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1546

A DNA sequence (GBSx1637) was identified in *S.agalactiae* <SEQ ID 4767> which encodes the amino acid sequence <SEQ ID 4768>. This protein is predicted to be transcriptional regulator, biotin repressor family. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2237(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14749 GB:Z99118 yrxA [Bacillus subtilis]
 Identities = 72/165 (43%), Positives = 112/165 (67%), Gaps = 2/165 (1%)

Query: 6   RRENILTTLKGTKEAISASTLAKIFSVSRQVIVGDIALLRAQQCDIISTPKGYL-MSSAL 64
           RR+ +L  LK +K  ++   LAK  +VSRQVIV DI+LL+A+   II+T +GY+ M +A
Sbjct: 12  RRDQLLLWLKESKSPLTGGELAKKANVSRQVIVQDISLLKAKNVPIIATSQGYVYMDAAA 71

Query: 65  STHQFTARLV-CQHGIEQTEEELEIILRYQGIIMNVEVEHPIYGMLTAPLNIQSQKDIDN 123
             HQ   R++ C HG E+TEEEL++I+     + +V++EHP+YG LTA + + ++K++ +
Sbjct: 72  QQHQQAERIIACLHGPERTEEELQLIVDEGVTVKDVKIEHPVYGDLTAAIQVGTRKEVSH 131

Query: 124 FTAKLKVSNAELLSSLTDGLHTHMISCQDQSVFDQICEALKKAGI 168
           F  K+  +NA  LS LTDG+H H ++  D+   DQ C+AL++AGI
Sbjct: 132 FIKKINSTNAAYLSQLTDGVHLHTLTAPDEHRIDQACQALEEAGI 176
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4769> which encodes the amino acid sequence <SEQ ID 4770>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2971(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 109/170 (64%), Positives = 136/170 (79%)

Query: 1   MKAQERRENILTTLKGTKEAISASTLAKIFSVSRQVIVGDIALLRAQQCDIISTPKGYLM 60
           MKA++RR+ I+  L   ++A+SA+ L K+  VSRQVIVGDIALLRAQQ DIISTPKGY+M
Sbjct: 1   MKAEDRRQKIIECLNSEQKAVSATRLGKLLGVSRQVIVGDIALLRAQQIDIISTPKGYIM 60

Query: 61  SSALSTHQFTARLVCQHGIEQTEEELEIILRYQGIIMNVEVEHPIYGMLTAPLNIQSQKD 120
```

```
            S+AL +HQF AR+VCQH +E+T++ELEIIL +QGII  VEVEHPIYGM+TAPLNI++  D
Sbjct: 61  STALYSHQFCARIVCQHNVEETKKELEIILAHQGIITTVEVEHPIYGMITAPLNIKTHSD 120

Query: 121 IDNFTAKLKVSNAELLSSLTDGLHTHMISCQDQSVFDQICEALKKAGILY 170
            + NF +KL  S AELLSSLT+GLH+H+ISC  Q  F  I   L+ AGILY
Sbjct: 121 VTNFMSKLSQSKAELLSSLTEGLHSHLISCPSQEAFLAIKHDLELAGILY 170
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1547

A DNA sequence (GBSx1638) was identified in *S.agalactiae* <SEQ ID 4771> which encodes the amino acid sequence <SEQ ID 4772>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.44   Transmembrane  143 - 159 ( 138 - 165)
   INTEGRAL    Likelihood = -8.17   Transmembrane  164 - 180 ( 160 - 184)
   INTEGRAL    Likelihood = -7.17   Transmembrane   56 -  72 (  53 -  78)
   INTEGRAL    Likelihood = -5.63   Transmembrane   24 -  40 (  21 -  44)
   INTEGRAL    Likelihood = -4.94   Transmembrane  113 - 129 ( 108 - 131)
   INTEGRAL    Likelihood = -2.39   Transmembrane   86 - 102 (  86 - 103)
   INTEGRAL    Likelihood = -1.06   Transmembrane  203 - 219 ( 203 - 219)

----- Final Results -----
            bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10069> which encodes amino acid sequence <SEQ ID 10070> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC18360 GB:AF064763 putative membrane spanning protein
         [Lactococcus lactis subsp. cremoris]
 Identities = 97/188 (51%), Positives = 133/188 (70%)

Query: 38  IMLYMFPQNMIAIMQKMPGLYFGAIILELVLVFVASGAARRNTPAALPLFLIYSALNGFT 97
            IM+  F  NM AI+Q           I+ LV+V    G A +N+  ALP+F+ Y+A  GF
Sbjct: 1   IMITFFLDNMRAILQSGSLFLLVLWIIPLVMVVSLQGLAMKNSKMALPIFIGYAAFMGFL 60

Query: 98  LSFIIARYTQTTVLQAFITSAAVFFAMALIGAKTKKDLSGMRKALMAALIGILIASLVNL 157
            +SF +  YT T +  AFIT++A+FF +++ G  TK++LSGM KAL  A+ G+++A L+NL
Sbjct: 61  ISFTLLMYTATDITLAFITASAMFFGLSVYGRFTKRNLSGMGKALGVAVWGLIVAMLLNL 120

Query: 158 FIGSGGMSYIISIVCVIIFSGLIAYDNQMIKYVYNSQGGQVADGWAVSMALSLYLDFINL 217
            F  S G++ +IS+V V+IFSGLIA+DNQ I  VYN+  GQV+DGWA+SMALSLYLDFIN+
Sbjct: 121 FFASTGLTILISLVGVVIFSGLIAWDNQKITQVYNAHNGQVSDGWAISMALSLYLDFINM 180

Query: 218 FLNILRLF 225
            FL +LRLF
Sbjct: 181 FLFLLRLF 188
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4773> which encodes the amino acid sequence <SEQ ID 4774>. Analysis of this protein sequence reveals the following:

```
     Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.97   Transmembrane  143 - 159 ( 138 - 165)
   INTEGRAL    Likelihood = -5.89   Transmembrane  164 - 180 ( 160 - 184)
   INTEGRAL    Likelihood = -5.68   Transmembrane   56 -  72 (  55 -  77)
   INTEGRAL    Likelihood = -4.78   Transmembrane  113 - 129 ( 110 - 130)
```

```
INTEGRAL    Likelihood = -2.81   Transmembrane  203 - 219 ( 203 - 222)
INTEGRAL    Likelihood = -2.76   Transmembrane   24 -  40 (  23 -  41)
INTEGRAL    Likelihood = -2.76   Transmembrane   86 - 102 (  86 - 104)

----- Final Results -----
           bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC18360 GB:AF064763 putative membrane spanning protein
           [Lactococcus lactis subsp. cremoris]
 Identities = 90/189 (47%), Positives = 133/189 (69%)

Query: 38  LMLYPFRENLISILVNQPMIYYGAAIIELILVFVASSAARKNTPAALPIFLIYSALNGFT 97
            +M+  F +N+ +IL +  +      II L++V      A KN+  ALPIF+ Y+A  GF
Sbjct: 1   IMITFFLDNMRAILQSGSLFLLVLWIIPLVMVVSLQGLAMKNSKMALPIFIGYAAFMGFL 60

Query: 98  LSFIIVAYAQTTVFQAFLSSAAVFFAMSIIGVKTKRDMSGLRKAMFAALIGVVVASLINL 157
            +SF ++ Y  T +  AF++++A+FF +S+ G  TKR++SG+ KA+  A+ G++VA L+NL
Sbjct: 61  ISFTLLMYTATDITLAFITASAMFFGLSVYGRFTKRNLSGMGKALGVAVWGLIVAMLLNL 120

Query: 158 FIGSGMMSYVISVISVLIFSGLIASDNQMIKRVYQATNGQVGDGWAVAMALSLYLDFINL 217
            F  S  ++ +IS++ V+IFSGLIA DNQ I +VY A NGQV DGWA++MALSLYLDFIN+
Sbjct: 121 FFASTGLTILISLVGVVIFSGLIAWDNQKITQVYNAHNGQVSDGWAISMALSLYLDFINM 180

Query: 218 FISLLRIFG 226
            F+ LLR+FG
Sbjct: 181 FLFLLRLFG 189
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 167/229 (72%), Positives = 202/229 (87%)

Query: 1   MNDNVIYTQSDSGLNQFFAKIYGLVGIGVGLSAAVSAIMLYMFPQNMIAIMQKMPGLYFG 60
            MND+VIYTQSD GLNQFFAKIY LVG+GVGLSA VS +MLY F +N+I+I+   P +Y+G
Sbjct: 1   MNDHVIYTQSDVGLNQFFAKIYSLVGMGVGLSAFVSYLMLYPFRENLISILVNQPMIYYG 60

Query: 61  AIILELVLVFVASGAARRNTPAALPLFLIYSALNGFTLSFIIARYTQTTVLQAFITSAAV 120
            A I+EL+LVFVAS AAR+NTPAALP+FLIYSALNGFTLSFII  Y QTTV QAF++SAAV
Sbjct: 61  AAIIELILVFVASSAARKNTPAALPIFLIYSALNGFTLSFIIVAYAQTTVFQAFLSSAAV 120

Query: 121 FFAMALIGAKTKKDLSGMRKALMAALIGILIASLVNLFIGSGGMSYIISIVCVIIFSGLI 180
            FFAM++IG KTK+D+SG+RKA+ AALIG+++ASL+NLFIGSG MSY+IS++ V+IFSGLI
Sbjct: 121 FFAMSIIGVKTKRDMSGLRKAMFAALIGVVVASLINLFIGSGMMSYVISVISVLIFSGLI 180

Query: 181 AYDNQMIKYVYNSQGGQVADGWAVSMALSLYLDFINLFLNILRLFARND 229
            A DNQMIK VY +   GQV DGWAV+MALSLYLDFINLF+++LR+F RND
Sbjct: 181 ASDNQMIKRVYQATNGQVGDGWAVAMALSLYLDFINLFISLLRIFGRND 229
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1548

A DNA sequence (GBSx1639) was identified in *S.agalactiae* <SEQ ID 4775> which encodes the amino acid sequence <SEQ ID 4776>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2495(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10071> which encodes amino acid sequence <SEQ ID 10072> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4777> which encodes the amino acid sequence <SEQ ID 4778>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3277(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 127/163 (77%), Positives = 141/163 (85%)

Query: 7    YQDDKDFMDLVGHLIDHPRFQKLEAIVQHHHSTRLEHSINVSYTSYKIAKKFGWDASSTA 66
            Y +DK++M+ VGHLI HPRFQKL  IVQH HSTRLEHSINVSY+SYK+AK+FGWDA STA
Sbjct: 3    YTEDKEYMEHVGHLIAHPRFQKLSHIVQHQHSTRLEHSINVSYSSYKLAKRFGWDAKSTA 62

Query: 67   RGGLLHDFFYYDWRVTKFNKSHAWVHPRIAVRNARKLTDLNAREEDIILKHMWGATIAPP 126
            RGGLLHDFFYYDWRVTKFNK HAWVHPRIAVRNA+KLT+LN +EEDIILKHMWGATIA P
Sbjct: 63   RGGLLHDFFYYDWRVTKFNKGHAWVHPRIAVRNAKKLTELNKKEEDIILKHMWGATIAFP 122

Query: 127  RYKESYIVTMVDKYWAVREASRPLKRIFKKPIRFSRKFLGSHN 169
            RYKESYIVTMVDKYWAV+EA  PL++ +       RK L SHN
Sbjct: 123  RYKESYIVTMVDKYWAVKEAVTPLRQKWSNRRFLRRKTLQSHN 165
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1549

A DNA sequence (GBSx1640) was identified in *S.agalactiae* <SEQ ID 4779> which encodes the amino acid sequence <SEQ ID 4780>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.03   Transmembrane  213 - 229 ( 212 - 229)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9413> which encodes amino acid sequence <SEQ ID 9414> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14825 GB:Z99118 similar to rRNA methylase [Bacillus subtilis]
 Identities = 96/228 (42%), Positives = 143/228 (62%), Gaps = 5/228 (2%)

Query: 3    QKKYRKSSYLIEGWHLFEEAEKYGAQFLNIFVT-ETAIDR-LRKPERAIVVTDDVLKELT 60
            +++ + +++LIEG HL EEA K       I V  ET I   L    +  ++++D    +T
Sbjct: 22   KERTKTNTFLIEGEHLVEEALKSPGIVKEILVKDETRIPSDLETGIQCYMLSEDAFSAVT 81

Query: 61   DSQTPQGIVAEIAFQETRWTDIKKGRFLVLEDVQDPGNLGTMVRTADAANFDAVFLSQKS 120
```

```
           +++TPQ I A     E +    +K   L+++ VQDPGNLGTM+RTADAA  DAV L   +
Sbjct: 82  ETETPQQIAAVCHMPEEKLATARK--VLLIDAVQDPGNLGTMIRTADAAGLDAVVLGDGT 139

Query: 121 ADLYNQKTLRSMQGSHFHLPVFRVEIEQFVNFCKAEGITMIATTLSEQSVNYKNLPKYDY 180
           AD +N KTLRS QGSHFH+PV R  +  +V+  KAEG+ +  T L +    Y+ +P+ +
Sbjct: 140 ADAFNGKTLRSAQGSHFHIPVVRRNLPSYVDELKAEGVKVYGTAL-QNGAPYQEIPQSES 198

Query: 181 FALIMGNEGQGISKTMTEEADVLAHIEMPGQAESLNVAVAAGVVIFSL 228
           FALI+GNEG G+   + E+ D+  ++ + GQAESLNVAVAA ++++ L
Sbjct: 199 FALIVGNEGAGVDAALLEKTDLNLYVPLYGQAESLNVAVAAAILVYHL 246
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4781> which encodes the amino acid sequence <SEQ ID 4782>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.97   Transmembrane  229 - 245 ( 228 - 245)

----- Final Results -----
               bacterial membrane --- Certainty=0.2190(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 141/229 (61%), Positives = 178/229 (77%)

Query: 1   MLQKKYRKSSYLIEGWHLFEEAEKYGAQFLNIFVTETAIDRLRKPERAIVVTDDVLKELT 60
           +LQKK+RK SYLIEGWHLFEEA+K G  F +IFV E  ++RL   +  ++V+  VLKELT
Sbjct: 17  LLQKKHRKQSYLIEGWHLFEEAQKSGQVFRHIFVLEEMVERLAGEQELVIVSPQVLKELT 76

Query: 61  DSQTPQGIVAEIAFQETRWTDIKKGRFLVLEDVQDPGNLGTMVRTADAANFDAVFLSQKS 120
           DS +PQGIVAE+   +  +    KG++LVLEDVQDPGNLGT++RTADAA FD VFLS+KS
Sbjct: 77  DSPSPQGIVAEVEIPKLAFPSDYKGKYLVLEDVQDPGNLGTIIRTADAARFDGVFLSEKS 136

Query: 121 ADLYNQKTLRSMQGSHFHLPVFRVEIEQFVNFCKAEGITMIATTLSEQSVNYKNLPKYDY 180
           AD+YNQKTLRSMQGSHFHLP++R ++ Q     +     ++ATTLS++SV+YK+L  ++
Sbjct: 137 ADIYNQKTLRSMQGSHFHLPIWRTDVYQLCRELQEYETPILATTLSKKSVDYKSLTHHER 196

Query: 181 FALIMGNEGQGISKTMTEEADVLAHIEMPGQAESLNVAVAAGVVIFSLI 229
            AL++GNEGQGIS  M   AD L HI MPGQAESLNVAVAAG++IFSLI
Sbjct: 197 LALVLGNEGQGISAEMAALADQLVHITMPGQAESLNVAVAAGILIFSLI 245
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8839> and protein <SEQ ID 8840> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     -7.98
GvH: Signal Score (-7.5): -3.86
     Possible site: 37
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -3.03 threshold:  0.0
   INTEGRAL    Likelihood = -3.03   Transmembrane  213 - 229 ( 212 - 229)
   PERIPHERAL  Likelihood =  5.14     149
 modified ALOM score:   1.11

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02468(259 - 984 of 1287)
EGAD|107730|BS2859(4 - 246 of 248) hypothetical protein {Bacillus subtilis}
GP|1770029|emb|CAA99602.1||Z75208 hypothetical protein {Bacillus subtilis}
GP|2635330|emb|CAB14825.1||Z99118 similar to rRNA methylase {Bacillus subtilis}
PIR|G69984|G69984 rRNA methylase homolog ysgA - Bacillus subtilis
%Match = 20.3
%Identity = 43.0  %Similarity = 62.3
Matches = 105  Mismatches = 87  Conservative Sub.s = 47

186       216       246       276       306       330       360       390
A*RNPTP*TRPETIK*TFFIT*PLF*YNRXMTTIITSKSNNLIKKTKKLLQKKYR--KSSYLIEGWHLFEEAEKYGAQFL
                                  | |  |  :|  |||  || |   :::|||| || ||| |
                               MKQIESAKNQKVKDWKKLHTKKERTKTNTFLIEGEHLVEEALKSPGIVK
                                       10        20        30        40

417       444       474       504       534       564       594       624
NIFVT-ETAI-DRLRKPERAIVVTDDVLKELTDSQTPQGIVAEIAFQETRWTDIKKGRFLVLEDVQDPGNLGTMVRTADA
 |:|  || |   |    :  ::::| :  :|:::||| | |     | :       : |::: ||||||||||:|||||
EILVKDETRIPSDLETGIQCYMLSEDAFSAVTETETPQQIAAVCHMPEEKLA--TARKVLLIDAVQDPGNLGTMIRTADA
         60        70        80        90        100        110        120

654       684       714       744       774       804       834       864
ANFDAVFLSQKSADLYNQKTLRSMQGSHFHLPVFRVEIEQFVNFCKAEGITMIATTLSEQSVNYKNLPKYDYFALIMGNE
| :||| |   :|| :| ||||| ||||||:|| |  :  :|:  ||||: :  | | :    |: :|: : ||||:|||
AGLDAVVLGDGTADAFNGKTLRSAQGSHFHIPVVRRNLPSYVDELKAEGVKVYGTAL-QNGAPYQEIPQSESFALIVGNE
          140       150       160       170       180        190       200

894       924       954       984       1014      1044      1074      1104
GQGISKTMTEEADVLAHIEMPGQAESLNVAVAAGVVIFSLI*VHML*YPQRGDYNEKVSRR*GLHGFGRSPY*PSTFPKT
| |:   : |: |:  :: : |||||||||||| :::: |
GAGVDAALLEKTDLNLYVPLYGQAESLNVAVAAAILVYHLRG
           220       230       240
```

SEQ ID 8840 (GBS430) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 5; MW 29kDa).

GBS430-GST was purified as shown in Figure 220, lane 8.

Example 1550

A DNA sequence (GBSx1641) was identified in *S.agalactiae* <SEQ ID 4783> which encodes the amino acid sequence <SEQ ID 4784>. This protein is predicted to be acylphosphatase (acyP). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10073> which encodes amino acid sequence <SEQ ID 10074> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD36630 GB:AE001801 acylphosphatase, putative [Thermotoga maritima]
 Identities = 35/88 (39%), Positives = 52/88 (58%), Gaps = 3/88 (3%)

Query: 24  MKKVHLIVSGRVQGVGFRYATYSLALEIGDIYGRVWNNDDGTVEILAQSTDSNKMTQFIQ 83
            MK + + V G VQGVGFRY T  +A  +G + G V N DDG+V I A+  D N + +F+
```

```
Sbjct: 1    MKALKIRVEGIVQGVGFRYFTRRVAKSLG-VKGYVMNMDDGSVFIHAEG-DENALRRFLN 58

Query: 84   KIRKGPSKWSKVTYVDIKLDNFDDFNDF 111
            ++ KGP   + VT V ++    + + DF
Sbjct: 59   EVAKGPPA-AVVTNVSVEETTPEGYEDF 85
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4785> which encodes the amino acid sequence <SEQ ID 4786>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 69/95 (72%), Positives = 85/95 (88%)

Query: 19   KRGQVMKKVHLIVSGRVQGVGFRYATYSLALEIGDIYGRVWNNDDGTVEILAQSTDSNKM 78
            K   +M+KV LIVSGRVQGVGFRYAT++LAL+IGDIYGRVWNN+DGTVEILAQS DS+K+
Sbjct: 7    KEALLMQKVRLIVSGRVQGVGFRYATHTLALDIGDIYGRVWNNNDGTVEILAQSKDSDKI 66

Query: 79   TQFIQKIRKGPSKWSKVTYVDIKLDNFDDFNDFKM 113
              FIQ++RKGPSKW+KVTYVD+ + NF+DF DF++
Sbjct: 67   ATFIQEVRKGPSKWAKVTYVDVTMANFEDFQDFQI 101
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1551

A DNA sequence (GBSx1642) was identified in *S.agalactiae* <SEQ ID 4787> which encodes the amino acid sequence <SEQ ID 4788>. This protein is predicted to be membrane protein homolog (yidC). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> May be a lipoprotein
   INTEGRAL    Likelihood =-12.52   Transmembrane    60 -  76 (  54 -  83)
   INTEGRAL    Likelihood = -3.66   Transmembrane   178 - 194 ( 177 - 196)
   INTEGRAL    Likelihood = -2.76   Transmembrane   140 - 156 ( 137 - 157)
   INTEGRAL    Likelihood = -2.60   Transmembrane   216 - 232 ( 213 - 232)

----- Final Results -----
               bacterial membrane --- Certainty=0.6010(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10075> which encodes amino acid sequence <SEQ ID 10076> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF03934 GB:AF139908 membrane protein homolog [Listeria
           monocytogenes]
 Identities = 82/222 (36%), Positives = 133/222 (58%), Gaps = 4/222 (1%)

Query: 44   PMANLITYFAQHQGLGFGVAIIIVTVIVRVVILPLGLYQSWKASYQAEKMAYFKPLFEPI 103
            P  + I + A+  G  +G+AIII T+++R +I+PL L  +        KMA  KP  + I
Sbjct: 3    PFTSFIMFVAKFVGGNYGIAIIITTLLIRALIMPLNLRTAKAQMGMQSKMAVAKPEIDEI 62
```

```
Query: 104 NERLRNAKTQEEKLAAQTELMTAQRENGLSMFGGIGCLPLLIQMPFFSAIFFAARYTPGV 163
             RL+ A ++EE+   Q E+M    +   ++     +GCLPLLIQMP   A ++A R +  +
Sbjct: 63  QARLKRATSKEEQATIQKEMMAVYSKYNINPMQ-MGCLPLLIQMPILMAFYYAIRGSSEI 121

Query: 164 SSATFLGLNLGQKSLTLTVIIAILYFVQSWLSMQGVPDEQRQQMKTMMYLMPIMMVFMSI 223
           +S TFL  NLG   + L +I  ++Y  Q ++SM G   EQ++QMK +  + PIM++F+S
Sbjct: 122 ASHTFLWFNLGSPDMVLAIIAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSPIMILFVSF 181

Query: 224 SLPASVALYWFIGGIFSIIQQLVT--TYVLK-PKLRRKVEEE 262
           + P+++ALYW +GG+F   Q L+T   Y+ K P+++   +EE
Sbjct: 182 TAPSALALYWAVGGLFLAGQTLLTKKLYMNKHPEIKVMEQEE 223
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4789> which encodes the amino acid sequence <SEQ ID 4790>. Analysis of this protein sequence reveals the following:

```
        Possible site: 31

>>> May be a lipoprotein
   INTEGRAL    Likelihood = -9.55    Transmembrane   62 -  78 (  54 -  82)
   INTEGRAL    Likelihood = -2.81    Transmembrane  178 - 194 ( 177 - 195)
   INTEGRAL    Likelihood = -0.90    Transmembrane  216 - 232 ( 215 - 232)

----- Final Results -----
               bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF03934 GB:AF139908 membrane protein homolog [Listeria monocytogenes]
 Identities = 89/218 (40%), Positives = 132/218 (59%), Gaps = 2/218 (0%)

Query: 43  KPMSYFIDYFANNAGLGYGLAIIIVTIIVRTLILPLGLYQSWKASYQSEKMAFLKPVFEP 102
            +P + FI + A   G  YG+AIII T+++R LI+PL L  +        KMA  KP  +
Sbjct: 2   QPFTSFIMFVAKFVGGNYGIAIIITTLLIRALIMPLNLRTAKAQMGMQSKMAVAKPEIDE 61

Query: 103 INKRIKQANSQEEKMAAQTELMAAQRAHGINPLGGIGCLPLLIQMPFFSAMYFAAQYTKG 162
            I  R+K+A S+EE+   Q E+MA    + INP+  +GCLPLLIQMP   A Y+A + +
Sbjct: 62  IQARLKRATSKEEQATIQKEMMAVYSKYNINPMQ-MGCLPLLIQMPILMAFYYAIRGSSE 120

Query: 163 VSTSTFMGIDLGSRSLVLTAIIAALYFFQSWLSMMAVSEEQREQMKTMMYTMPIMMIFMS 222
            +++ TF+  +LGS  +VL  I   +Y  Q ++SM+  S EQ++QMK +    PIM++F+S
Sbjct: 121 IASHTFLWFNLGSPDMVLAIIAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSPIMILFVS 180

Query: 223 FSLPAGVGLYWLVGGFFSIIQQLITTYLLKPRLHKQIK 260
            F+ P+ + LYW VGG F   Q L+T  L   + H +IK
Sbjct: 181 FTAPSALALYWAVGGLFLAGQTLLTKKLYMNK-HPEIK 217
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 203/309 (65%), Positives = 254/309 (81%), Gaps = 2/309 (0%)

Query: 1   MKKTLKRILFSSLSLSMLLLLTGCVSVDKAGKPYGVIWNTLGVPMANLITYFAQHQGLGF 60
            +K TL RILFS L+LS+LL LTGCV  D  G P G+IW  LG PM+  I YFA + GLG+
Sbjct: 1   LKLTLNRILFSGLALSILLTLTGCVGRDAHGNPKGMIWEFLGKPMSYFIDYFANNAGLGY 60

Query: 61  GVAIIIVTVIVRVVILPLGLYQSWKASYQAEKMAYFKPLFEPINERLRNAKTQEEKLAAQ 120
            G+AIIIVT+IVR +ILPLGLYQSWKASYQ+EKMA+ KP+FEPIN+R++ A +QEEK+AAQ
Sbjct: 61  GLAIIIVTIIVRTLILPLGLYQSWKASYQSEKMAFLKPVFEPINKRIKQANSQEEKMAAQ 120

Query: 121 TELMTAQRENGLSMFGGIGCLPLLIQMPFFSAIFFAARYTPGVSSATFLGLNLGQKSLTL 180
            TELM AQR +G++  GGIGCLPLLIQMPFFSA++FAA+YT GVS++TF+G++LG +SL L
Sbjct: 121 TELMAAQRAHGINPLGGIGCLPLLIQMPFFSAMYFAAQYTKGVSTSTFMGIDLGSRSLVL 180

Query: 181 TVIIAILYFVQSWLSMQGVPDEQRQQMKTMMYLMPIMMVFMSISLPASVALYWFIGGIFS 240
            T IIA LYF QSWLSM   V +EQR+QMKTMMY MPIMM+FMS SLPA V LYW +GG FS
Sbjct: 181 TAIIAALYFFQSWLSMMAVSEEQREQMKTMMYTMPIMMIFMSFSLPAGVGLYWLVGGFFS 240
```

```
Query: 241 IIQQLVTTYVLKPKLRRKVEEEYTKNPPKAYKANNARKDVTNSTKATESNQAIITSKKTN 300
           IIQQL+TTY+LKP+L ++++EEY KNPPKAY++ ++RKDVT S    ++N  +    K+N
Sbjct: 241 IIQQLITTYLLKPRLHKQIKEEYAKNPPKAYQSTSSRKDVTPSQNMEQAN--LPKKIKSN 298

Query: 301 RNAGKQKRR 309
           RNAGKQ++R
Sbjct: 299 RNAGKQRKR 307
```

A related GBS gene <SEQ ID 8841> and protein <SEQ ID 8842> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 23   Crend: 6
McG: Discrim Score:      8.74
GvH: Signal Score (-7.5): -1.47
     Possible site: 16
>>> May be a lipoprotein
ALOM program   count: 4 value: -12.52 threshold:  0.0
   INTEGRAL    Likelihood =-12.52   Transmembrane   60 -  76 (  54 -  83)
   INTEGRAL    Likelihood = -3.66   Transmembrane  178 - 194 ( 177 - 196)
   INTEGRAL    Likelihood = -2.76   Transmembrane  140 - 156 ( 137 - 157)
   INTEGRAL    Likelihood = -2.60   Transmembrane  216 - 232 ( 213 - 232)
   PERIPHERAL  Likelihood =  0.74     235
 modified ALOM score:   3.00

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6010(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
37.9/63.7% over 193aa
                                                          Bacillus subtilis
   EGAD|45886| hypothetical 30.7 kd lipoprotein in glnq-ansr intergenic region precursor
Insert characterized
   SP|P54544|YQJG_BACSU HYPOTHETICAL 30.7 KDA LIPOPROTEIN IN GLNQ-ANSR INTERGENIC REGION
PRECURSOR. Insert characterized
  GP|1303958|dbj|BAA12613.1||D84432 YqjG Insert characterized
  GP|2634823|emb|CAB14320.1||Z99116    similar   to    lipoprotein    SpoIIIJ-like    Insert
characterized
  PIR|G69963|G69963 lipoprotein SpoIIIJ-like homolog yqjG - Insert characterized

ORF02470(478 - 1038 of 1530)
EGAD|45886|BS2384(63 - 256 of 275) hypothetical 30.7 kd lipoprotein in glnq-ansr intergenic
region precursor {Bacillus subtilis}SP|P54544|YQJG_BACSU HYPOTHETICAL 30.7 KDA LIPOPROTEIN
IN GLNQ-ANSR INTERGENIC REGION PRECURSOR.GP|1303958|dbj|BAA12613.1||D84432 YqjG {Bacillus
subtilis}GP|2634823|emb|CAB14320.1||Z99116 similar to lipoprotein SpoIIIJ-like {Bacillus
subtilis}PIR|G69963|G69963 lipoprotein SpoIIIJ-like homolog yqjG - Bacillus subtilis
%Match = 13.0
%Identity = 37.9  %Similarity = 63.7
Matches = 72  Mismatches = 65  Conservative Sub.s = 49

252       282       312       342       372       402       432       462
FCGSIV*FLKKK*NR*VY*KLEELKTLKKTLKRILFSSLSLSMLLLLTGCVSVDKAGKPYGVIWNTLGVPMANLITYFAQ

                       MLKTYQKLLAMGIFLIVLCSGNAAFAATNQVGGLSNVGFFHDYLIEPFSALLKGVAG
                               10        20        30        40        50

492       522       552       582       612       642       672       702
HQGLGFGVAIIIVTVIVRVVILPLGLYQSWKASYQAEKMAYFKPLFEPINERLRNAKTQEEKLAAQTELMTAQRENGLSM
     :|::||:||:|||:|:||| : |  |     ||||  ||  : |  :|:  |  |::   | |:|   :|: ::
LFHGEYGLSIILVTIIVRIVVLPLFVNQFKKQRIFQEKMAVIKPQVDSIQVKLKKTKDPEKQKELQMEMMKLYQEHNINP
           70        80        90       100       110       120       130
```

```
732       762       792       822       852                 894       918
FGGIGCLPLLIQMPFFSAIFFAARYTPGVSSATFLGLNLGQKSLTLTVIIAILYFVQSW----LSMQ--GVPDE--QRQQ
   :||||:||| |    :::| | || ::| :|| ::|||  : :::   |:||||::    || :   ||     :|
-LAMGCLPMLIQSPIMIGLYYAIRSTPEIASHSFLWFSLGQSDILMSLSAGIMYFVQAYIAQKLSAKYSAVPQNPAAQQS
           150       160       170       180       190       200       210

948       978      1008      1038      1068      1098      1128      1158
MKTMMYLMPIMMVFMSISLPASVALYWFIGGIFSIIQQLVTTYVLKPKLRRKVEEEYTKNPPKAYKANNARKDVTNSTKA
 | |::: |:||   |:::||:: ||||  |:|  :| :|         :     |  |
AKLMVFIFPVMMTIFSLNVPAALPLYWFTSGLFLTVQNIVLQMTHHKSKKTAALTESVK
           230       240       250       260       270

37.2/62.0% over 220aa
Listeria monocytogenes
  GP|6117974| membrane protein homolog Insert characterized

ORF02470(430 - 1086 of 1530)
GP|6117974|gb|AAF03934.1|AF139908_4|AF139908(3  -  223  of  237)  membrane  protein  homolog
{Listeria monocytogenes}
%Match = 14.6
%Identity = 37.1  %Similarity = 62.0
Matches = 82  Mismatches = 81  Conservative Sub.s = 55

285       315       345       375       405       435       465       495
K*NR*VY*KLEELKTLKKTLKRILFSSLSLSMLLLLTGCVSVDKAGKPYGVIWNTLGVPMANLITYFAQHQGLGFGVAII
                                                          |  ::| : |:  |  :|:|||
                                                        IQPFTSFIMFVAKFVGGNYGIAII
                                                                10        20

525       555       585       615       645       675       705       735
IVTVIVRVVILPLGLYQSWKASYQAEKMAYFKPLFEPINERLRNAKTQEEKLAAQTELMTAQRENGLSMFGGIGCLPLLI
| |:::| :|:|| |  :        |||  ||  : |  ||: | ::||:   | |:|    :  ::    :|||||||
ITTLLIRALIMPLNLRTAKAQMGMQSKMAVAKPEIDEIQARLKRATSKEEQATIQKEMMAVYSKYNINP-MQMGCLPLLI
              40        50        60        70        80        90       100

765       795       825       855       885       915       945       975
QMPFFSAIFFAARYTPGVSSATFLGLNLGQKSLTLTVIIAILYFVQSWLSMQGVPDEQRQQMKTMMYLMPIMMVFMSISL
||| : | ::| | :  ::| ||| :|||   : | :|  ::|: | ::|| |   ||::||| :  : |||::|:| :
QMPILMAFYYAIRGSSEIASHTFLWFNLGSPDMVLAIIAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSPIMILFVSFTA
              120       130       140       150       160       170       180

1005      1035                1086      1116      1146      1176      1206
PASVALYWFIGGIFSIIQQLVTT--YVLK-PKLRRKVEEEYTKNPPKAYKANNARKDVTNSTKATESNQAIITSKKTNRN
|:::|||| :||:|   | |:|   |: | |:::   :||
PSALALYWAVGGLFLAGQTLLTKKLYMNKHPEIKVMEQEEKEFEQIVEEQKKEK
              200       210       220       230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1552

A DNA sequence (GBSx1644) was identified in *S.agalactiae* <SEQ ID 4791> which encodes the amino acid sequence <SEQ ID 4792>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.98   Transmembrane   32 -  48 (  23 -  53)
   INTEGRAL    Likelihood = -9.18   Transmembrane  195 - 211 ( 189 - 213)
   INTEGRAL    Likelihood = -8.70   Transmembrane   72 -  88 (  62 -  93)

----- Final Results -----
               bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter
           (permease) [Bacillus subtilis]
 Identities = 116/217 (53%), Positives = 168/217 (76%)

Query: 2   INWDAIFNLELAVKAFPSVIQGLPYTIGLSLVGFILGAIVGFFVALMKMSHFRLLRYLAN 61
           I W+ IFN +LA+++FP VI+G+ YT+ +S V    G ++G F++L +MS   LLR+ A
Sbjct: 5   IQWEYIFNTKLAIESFPYVIKGIGYTLLISFVSMFAGTVIGLFISLARMSKLALLRWPAK 64

Query: 62  IHISLMRGIPLMVLLFLIYFGLPFIGIQLDAVTASIVGFTMMSSAYISEIIRAALLAVDH 121
           ++IS MRG+P++V+LF++YFG P+IGI+  AVTA+++GF++ S+AYI+EI R+A+ +V+
Sbjct: 65  LYISFMRGVPILVILFILYFGFPYIGIEFSAVTAALIGFSLNSAAYIAEINRSAISSVEK 124

Query: 122 GQWEAARALGLKTPTIYRGIIIPQATRIALPSLSNVLLDMVKSSSLTAMITVPDIFNNAK 181
           GQWEAA +LGL      RGII+PQ+ RIALP L+NVLLD++K+SSL AMITVP++  +AK
Sbjct: 125 GQWEAASSLGLSYWQTMRGIILPQSIRIALPPLANVLLDLIKASSLAAMITVPELLQHAK 184

Query: 182 IVGGTYSDYMTAYILVALIYWVICTLYAIIQDWWEKR 218
           I+GG   DYMT YIL ALIYW IC++ A+ Q+  EK+
Sbjct: 185 IIGGREFDYMTMYILTALIYWAICSIAAVFQNILEKK 221
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4793> which encodes the amino acid sequence <SEQ ID 4794>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.79    Transmembrane  186 - 202 ( 184 - 205)
   INTEGRAL    Likelihood = -5.84    Transmembrane   26 -  42 (  21 -  43)
   INTEGRAL    Likelihood = -4.78    Transmembrane   57 -  73 (  56 -  84)
   INTEGRAL    Likelihood = -1.59    Transmembrane   86 - 102 (  86 - 103)

----- Final Results -----
              bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter
           (permease) [Bacillus subtilis]
 Identities = 113/214 (52%), Positives = 157/214 (72%)

Query: 1   MINIPLMKDSLGFVLSGLPYTLGISLLSFFTGLFLGLGLALLGRSRQPLIHYLVRAYISI 60
           + N  L  +S  +V+ G+ YTL IS +S F G  +GL ++L   S+  L+ +  + YIS
Sbjct: 10  IFNTKLAIESFPYVIKGIGYTLLISFVSMFAGTVIGLFISLARMSKLALLRWPAKLYISF 69

Query: 61  MRGVPMIVVLFVLYFGLPYYGLELPALLCAYLGFSMVSAAYISEVFRSSIEAIDKGQWEA 120
           MRGVP++V+LF+LYFG PY G+E  A+  A +GFS+ SAAYI+E+ RS+I +++KGQWEA
Sbjct: 70  MRGVPILVILFILYFGFPYIGIEFSAVTAALIGFSLNSAAYIAEINRSAISSVEKGQWEA 129

Query: 121 AKALGLPYALMVKKIILPQAFRIAVPPLGNVIIDMVKSSSLAAMITVPDIFQNAKIIGGR 180
           A +LGL Y   ++ IILPQ+ RIA+PPL NV++D++K+SSLAAMITVP++ Q+AKIIGGR
Sbjct: 130 ASSLGLSYWQTMRGIILPQSIRIALPPLANVLLDLIKASSLAAMITVPELLQHAKIIGGR 189

Query: 181 EWDYMSMYILVAFIYWLIAFLLERYQEFLENKLA 214
           E+DYM+MYIL A IYW I  +   +Q  LE K A
Sbjct: 190 EFDYMTMYILTALIYWAICSIAAVFQNILEKKYA 223
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 110/213 (51%), Positives = 156/213 (72%)

Query: 7   IFNLELAVKAFPSVIQGLPYTIGLSLVGFILGAIVGFFVALMKMSHFRLLRYLANIHISL 66
           + N+ L   +   V+ GLPYT+G+SL+ F  G  +G  +AL+  S   L+ YL   +IS+
Sbjct: 1   MINIPLMKDSLGFVLSGLPYTLGISLLSFFTGLFLGLGLALLGRSRQPLIHYLVRAYISI 60
```

```
Query: 67  MRGIPLMVLLFLIYFGLPFIGIQLDAVTASIVGFTMMSSAYISEIIRAALLAVDHGQWEA 126
           MRG+P++V+LF++YFGLP+ G++L A+  + +GF+M+S+AYISE+ R+++ A+D GQWEA
Sbjct: 61  MRGVPMIVVLFVLYFGLPYYGLELPALLCAYLGFSMVSAAYISEVFRSSIEAIDKGQWEA 120

Query: 127 ARALGLKTPTIYRGIIIPQATRIALPSLSNVLLDMVKSSSLTAMITVPDIFNNAKIVGGT 186
           A+ALGL    + + II+PQA RIA+P L NV++DMVKSSSL AMITVPDIF NAKI+GG
Sbjct: 121 AKALGLPYALMVKKIILPQAFRIAVPPLGNVIIDMVKSSSLAAMITVPDIFQNAKIIGGR 180

Query: 187 YSDYMTAYILVALIYWVICTLYAIIQDWWEKRL 219
             DYM+ YILVA IYW+I  L    Q++ E +L
Sbjct: 181 EWDYMSMYILVAFIYWLIAFLLERYQEFLENKL 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1553

A DNA sequence (GBSx1645) was identified in *S.agalactiae* <SEQ ID 4795> which encodes the amino acid sequence <SEQ ID 4796>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12132 GB:Z99105 similar to amino acid ABC transporter
          (binding protein) [Bacillus subtilis]
 Identities = 127/276 (46%), Positives = 183/276 (66%), Gaps = 12/276 (4%)

Query: 3   KTILLGLVGLSAMTLAACS--NGQSSKETTWDNIKKDGVLKVATPATLYPTSYYDDHK-- 58
           K ++        + LAACS  N   SK+T W+ IK  G + VAT  TLYPTSY+D
Sbjct: 8   KAVIFSFTMAFFLILAACSGKNEADSKDTGWEQIKDKGKIVVATSGTLYPTSYHDTDSGS 67

Query: 59  -KLTGYEIDMMKAIAKKLKIKVKFVEVGVAESFTSVDSGKVDVAVNNFDTTPERLKKYNF 117
            KLTGYE+++++  AK+L +KV+F E+G+    T+V+SG+VD A N+ D T +R +K+ F
Sbjct: 68  DKLTGYEVEVVREAAKRLGLKVEFKEMGIDGMLTAVNSGQVDAAANDIDVTKDREEKFAF 127

Query: 118 SQPYKYSVGGMIVRADGSSKITAKDLSDWKGKKAGGGAGTQYMKIAKQQGAEPVIYDNVT 177
           S PYKYS G  IVR D  S I  K L D KGKKA G A T YM++A++ GA+ VIYDN T
Sbjct: 128 STPYKYSYGTAIVRKDDLSGI--KTLKDLKGKKAAGAATTVYMEVARKYGAKEVIYDNAT 185

Query: 178 NDVYLRDVSTGRTDFIPNDYYTQVIAVKYVTKQYPDIKVKM-GDVKYNPTEQGIVMSKKD 236
           N+ YL+DV+ GRTD I NDYY Q +A+      +PD+ + +  D+KY P +Q +VM K +
Sbjct: 186 NEQYLKDVANGRTDVILNDYYLQTLAL----AAFPDLNITIHPDIKYMPNKQALVMKKSN 241

Query: 237 KSLKTKIDAAIKDMKKDGSLKKISEKYYAGQDLTKE 272
            +L+ K++ A+K+M KDGSL K+S++++   D++K+
Sbjct: 242 AALQKKMNEALKEMSKDGSLTKLSKQFFNKADVSKK 277
```

There is also homology to SEQ ID 1190.

SEQ ID 4796 (GBS183) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 2; MW 33kDa).

GBS183-His was purified as shown in Figure 199, lane 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1554

A DNA sequence (GBSx1646) was identified in *S.agalactiae* <SEQ ID 4797> which encodes the amino acid sequence <SEQ ID 4798>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1514(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF09821 GB:AE001885 6-aminohexanoate-cyclic-dimer hydrolase
           [Deinococcus radiodurans]
 Identities = 178/488 (36%), Positives = 265/488 (53%), Gaps = 17/488 (3%)

Query: 5   DATAMVQAIKQHKISSQELVEQAIYKIEEQNVSVNAVVSKQYNEARQAAKYANESNA--- 61
           DA   +Q  ++ ++S++++   AI++ +  NV++NAVV   Y++    A+  + + A
Sbjct: 54  DALDLAQLFRRGELSAEDMCTAAIHRAQVVNVALNAVVYPLYDQGLAQARATDAARARGE 113

Query: 62  ----PFAGVPILLKDLGQNQKGQLSTSGSQLFKHYHAKQTDYLVQSFEKLGFIILGRTNT 117
               PFAGVP L+KD G    G   T G++ ++    +  D LV+ ++  G + LG+TNT
Sbjct: 114 QATGPFAGVPFLVKDFGSRLAGVPHTGGTRAYRDQIPEWDDELVRRWQAAGLLPLGKTNT 173

Query: 118 PEFGFKNISDGQLHGNVNLPFDHSRNAGGSSGGAAAAVSSGMVPIAGASDGGGSIRIPAS 177
           PEF    +++ +LHG    P+D  R  GGSSGG+A+AV++G+VP+AGA DGGGSIRIPAS
Sbjct: 174 PEFALMGVTEPELHGPTRNPWDLGRTPGGSSGGSASAVAAGIVPLAGAGDGGGSIRIPAS 233

Query: 178 FNGLIGLKPSRGRIPVGPSSYRGWQGASSHFALTKSVRDTKRLLYYLQSYQVES----PF 233
             GL GLKPSRGR+P G      WQGA+    LT+SVRD+  LL   Q     +    P
Sbjct: 234 CCGLFGLKPSRGRVPCGDGVGEPWQGAAVEHVLTRSVRDSAALLDLEQGPDAGAALFLPS 293

Query: 234 PLKKLSKESLFEFSVSKPLKIAVLMDSPLKTKVSSEAKAAIKEAADFLSQKGNHLELVEQ 293
           P +  S+E   E      L+I      PL   V  E  AA++ AA  L   G+ +E V
Sbjct: 294 PERPYSEEVGRE---PGRLRIGFSTAHPLGRSVHPECVAAVQGAARLLESLGHEVEEVAL 350

Query: 294 PLDGIHSMKTYCMMNSVETAAMFDDIEKSLGRSMEFSDMELMTWAMYQSGQRVLAKDYSK 353
           P DG    + + M+   ET A    +  +LGR    SD+E +TW + Q G+   A D++
Sbjct: 351 PWDGPALAQAFLMLYFGETGASLAALRDTLGRPARASDVEAVTWLLGQLGRSYSAADFAA 410

Query: 354 LLDSWDQFAATMARFHENYDLILTAATNQPAPFHGQFD---LDETLQKQLRHMGEFSVSE 410
              SW+  A  M RFH+NYDL+LT     P    G+     +   L +  + M    +
Sbjct: 411 ARASWNVHARAMGRFHQNYDLLLTPVLATPPLQIGELQPRGVQAALLRAAQQMDVSGLLR 470

Query: 411 QQDLIWKMFEDSMAWTPFTHQPNLTGQPSLAIPTHLTKEGLPLGVQLTAAKGREDLLLAV 470
           +   +  +  D +   P+T   NLTGQP++++P H T +GLP+GVQ  A   RED+LL +
Sbjct: 471 RSGQVDALATDILEKMPYTQLANLTGQPAMSVPLHWTADGLPVGVQFVAPLAREDVLLRL 530

Query: 471 AELFEKEK 478
           A   E+ +
Sbjct: 531 AGQLEQAR 538
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4047> which encodes the amino acid sequence <SEQ ID 4048>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
```

```
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
         bacterial outside --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 277/484 (57%), Positives = 348/484 (71%), Gaps = 2/484 (0%)

Query: 1   MVFKDATAMVQAIKQHKISSQELVEQAIYKIEEQNVSVNAVVSKQYNEARQAAKYANESN 60
           M ++DATAM  A++  + +  ELV QAIYK ++ N ++NA+ S+++  A + AK  + S
Sbjct: 1   MTYQDATAMAIAVQTGQTTPLELVTQAIYKAKKLNPTLNAITSERFEAALEEAKQRDFSG 60

Query: 61  APFAGVPILLKDLGQNQKGQLSTSGSQLFKHYHAKQTDYLVQSFEKLGFIILGRTNTPEF 120
            PFAGVP+ LKDLGQ  KG  STSGS+LFK Y A +TD  V+  E LGFIILGR+NTPEF
Sbjct: 61  LPFAGVPLFLKDLGQELKGHSSTSGSRLFKEYQATKTDLFVKRLEALGFIILGRSNTPEF 120

Query: 121 GFKNISDGQLHGNVNLPFDHSRNAGGSSGGAAAAVSSGMVPIAGASDGGGSIRIPASFNG 180
           GFKNISD  LHG VNLP D++RNAGGSSGGAAA VSSG+  +A ASDGGGSIRIPASFNG
Sbjct: 121 GFKNISDSSLHGPVNLPRDNTRNAGGSSGGAAALVSSGISALATASDGGGSIRIPASFNG 180

Query: 181 LIGLKPSRGRIPVGPSSYRGWQGASSHFALTKSVRDTKRLLYYLQSYQVESPFPLKKLSK 240
           LIGLKPSRGR+PVGP SYR WQGAS HFALTKSVRDT+ LLYYLQ  Q+ESPFPL  L+K
Sbjct: 181 LIGLKPSRGRMPVGPGSYRSWQGASVHFALTKSVRDTRNLLYYLQMEQMESPFPLATLTK 240

Query: 241 ESLFEFSVSKPLKIAVLMDSPLKTKVSSEAKAAIKEAADFLSQKGNHL-ELVEQPLDGIH 299
           +S+++ S+ +PL IA        + VS +   A+++A  +L ++G+ L EL E P++
Sbjct: 241 DSIYQ-SLQRPLTIAFYQRLSDGSPVSLDTAKALRQAVTWLREQGHQLVELEEFPVNMTE 299

Query: 300 SMKTYCMMNSVETAAMFDDIEKSLGRSMEFSDMELMTWAMYQSGQRVLAKDYSKLLDSWD 359
            ++ Y +MNSVETAAMF DIE + GR M   DME MTWA+YQSG+ + A  YS++L  WD
Sbjct: 300 VIRHYYIMNSVETAAMFADIEDTFGRPMTKDDMETMTWAIYQSGKDIPAWRYSQVLQKWD 359

Query: 360 QFAATMARFHENYDLILTAATNQPAPFHGQFDLDETLQKQLRHMGEFSVSEQQDLIWKMF 419
            ++ATMA FHE YDL+LT  TN PAP HG+   D  L   L     FS  EQ +L+  MF
Sbjct: 360 TYSATMASFHETYDLLLTFTTNTPAPKHGELVPDSKLMANLAQAEIFSSEEQFNLVETMF 419

Query: 420 EDSMAWTPFTHQPNLTGQPSLAIPTHLTKEGLPLGVQLTAAKGREDLLLAVAELFEKEKQ 479
             S+A  P+T  PNLTGQP++++PT+ TKEGL +G+QL AAKGREDLLL +AE FE
Sbjct: 420 GKSLAINPYTALPNLTGQPAISLPTYETKEGLSMGIQLIAAKGREDLLLGIAEQFEAAGL 479

Query: 480 FKGP 483
            K P
Sbjct: 480 LKIP 483
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1555

A DNA sequence (GBSx1647) was identified in *S.agalactiae* <SEQ ID 4799> which encodes the amino acid sequence <SEQ ID 4800>. This protein is predicted to be transcription elongation factor (greA). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5003(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14674 GB:Z99117 transcription elongation factor [Bacillus subtilis]
 Identities = 86/154 (55%), Positives = 114/154 (73%), Gaps = 1/154 (0%)
```

```
Query: 3   EKTYPMTQVEKDQLEKELEELKLVRRPEVVERIKIARSYGDLSENSEYDAAKDEQAFVEG 62
           EK +PMT   K +LE+ELE LK V+R EVVERIKIARS+GDLSENSEYD+AK+EQAFVEG
Sbjct: 4   EKVFPMTAEGKQKLEQELEYLKTVKRKEVVERIKIARSFGDLSENSEYDSAKEEQAFVEG 63

Query: 63  QIQILETKIRYAEIIDSDAVAKDEVAIGKTVLVQEVGTNDKDTYHIVGAAGADIFSGKIS 122
           ++  LE  IR A+II+ D    + V +GKTV   E+   D+++Y IVG+A AD F GKIS
Sbjct: 64  RVTTLENMIRNAKIIEDDG-GSNVVGLGKTVTFVELPDGDEESYTIVGSAEADPFEGKIS 122

Query: 123 NESPIAHALIGKKTGDLATIESPAGSYQVEIISV 156
           N+SPIA +L+GKK  +  T+++P G   V+I+ +
Sbjct: 123 NDSPIAKSLLGKKVDEEVTVQTPGGEMLVKIVKI 156
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4801> which encodes the amino acid sequence <SEQ ID 4802>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4434(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 145/160 (90%), Positives = 149/160 (92%)

Query: 1   MAEKTYPMTQVEKDQLEKELEELKLVRRPEVVERIKIARSYGDLSENSEYDAAKDEQAFV 60
           MAEKTYPMT  EK+QLEKELEELKLVRRPE+VERIKIARSYGDLSENSEYDAAKDEQAFV
Sbjct: 17  MAEKTYPMTLTEKEQLEKELEELKLVRRPEIVERIKIARSYGDLSENSEYDAAKDEQAFV 76

Query: 61  EGQIQILETKIRYAEIIDSDAVAKDEVAIGKTVLVQEVGTNDKDTYHIVGAAGADIFSGK 120
           EGQI  LETKIRYAEIIDSDAVAKDEVAIGKTV+VQEVGT DKDTYHIVGAAGADIFSGK
Sbjct: 77  EGQISTLETKIRYAEIIDSDAVAKDEVAIGKTVIVQEVGTTDKDTYHIVGAAGADIFSGK 136

Query: 121 ISNESPIAHALIGKKTGDLATIESPAGSYQVEIISVEKTN 160
           ISNESPIA ALIGKKTGD   IESPA +Y VEIISVEKTN
Sbjct: 137 ISNESPIAQALIGKKTGDKVRIESPAATYDVEIISVEKTN 176
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1556

A DNA sequence (GBSx1648) was identified in *S.agalactiae* <SEQ ID 4803> which encodes the amino acid sequence <SEQ ID 4804>. This protein is predicted to be aminodeoxychorismate lyase-like protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.64    Transmembrane  238 - 254 ( 230 - 260)

----- Final Results -----
               bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein
          [Streptococcus thermophilus]
 Identities = 135/210 (64%), Positives = 171/210 (81%)
```

```
Query: 373 KTTSTPYKADDFLKLVQDETFIKKMVAKYPNLLGSLPDKSKAIYQLEGYLFPATYNYYKD 432
           K +ST  K  DFLKL++D+ FI KM AKYP LL +LP+ + A Y LEGYLFPATYN + D
Sbjct: 5   KHSSTGLKEKDFLKLMKDDAFITKMKAKYPTLLANLPNSTDAKYVLEGYLFPATYNIHDD 64

Query: 433 TTLEGLVEDMISTMNTKMAPYYNTIKAKNMSVNDVLTLSSLVEKEGSTDEDRRKIASVFY 492
           TT+E L E+M+ TM+T ++PYY TI + N +VN++LTL+SLVEKEG+TD+DR+ IASVFY
Sbjct: 65  TTVESLAEEMLFTMDTHLSPYYATILSSNHNVNEILTLASLVEKEGATDDDRKNIASVFY 124

Query: 493 NRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNIYTNTGLMPGPVDSPSI 552
           NRL++  ALQSNIA+LY +GKLG +T+L EDA I+T+I SPYN Y + GLMPGPVDSPS+
Sbjct: 125 NRLNSDMALQSNIAVLYVLGKLGQETTLKEDATIDTNIDSPYNDYVHKGLMPGPVDSPSL 184

Query: 553 SAIEATIKPASTDYLYFVADVKTGNVYYAK 582
           SAIEA I P+ST Y+YFVADV TGNVY+A+
Sbjct: 185 SAIEAVINPSSTKYMYFVADVSTGNVYFAE 214
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4805> which encodes the amino acid sequence <SEQ ID 4806>. Analysis of this protein sequence reveals the following:

```
      Possible site: 59

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.91   Transmembrane  161 - 177 ( 155 - 183)

----- Final Results -----
               bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein
          [Streptococcus thermophilus]
 Identities = 135/212 (63%), Positives = 161/212 (75%)

Query: 295 KTKKAKTPFNEKDFLDLVTDEAFIQDMVKRYPKLLATIPTKEKAIYRLEGYLFPATYNYY 354
           K K + T   EKDFL L+ D+AFI  M  +YP LLA +P    A Y LEGYLFPATYN +
Sbjct: 3   KGKHSSTGLKEKDFLKLMKDDAFITKMKAKYPTLLANLPNSTDAKYVLEGYLFPATYNIH 62

Query: 355 KETTMRELVEDMLAAMDATLVPYYDKIAASGKTVNEVLTLASLVEKEGSTDDDRRQIASV 414
            +TT+  L E+ML  MD  L PYY  I +S   VNE+LTLASLVEKEG+TDDDR+ IASV
Sbjct: 63  DDTTVESLAEEMLFTMDTHLSPYYATILSSNHNVNEILTLASLVEKEGATDDDRKNIASV 122

Query: 415 FYNRLNSGMALQSNIAILYAMGKLGEKTTLAEDATIDTTINSPYNIYTNTGLMPGPVASS 474
           FYNRLNS MALQSNIA+LY +GKLG++TTL EDATIDT I+SPYN Y + GLMPGPV S
Sbjct: 123 FYNRLNSDMALQSNIAVLYVLGKLGQETTLKEDATIDTNIDSPYNDYVHKGLMPGPVDSP 182

Query: 475 GVSAIEATLNPASTDYLYFVANVHTGEVYYAK 506
            +SAIEA +NP+ST Y+YFVA+V TG VY+A+
Sbjct: 183 SLSAIEAVINPSSTKYMYFVADVSTGNVYFAE 214
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 310/603 (51%), Positives = 403/603 (66%), Gaps = 86/603 (14%)

Query: 1   MTEFNDDQHSNHDQKSFKEQILAELEEANRLRKLREEELYQKEQEAKEAARRTAQLMADY 60
           +T+F D    +  Q+SFKEQILAELE+AN++RK +EEEL+
Sbjct: 3   LTDFKDKDQQDQ-QRSFKEQILAELEKANQIRKEKEEELF-------------------- 41

Query: 61  EAQRLKDEREARAKALETKQRLEEQEKARIEAKLLAEAAREEERRQAEQALASQEEQVIN 120
                             ++ LE +E AR  A+L AE  R++         A Q+E + +
Sbjct: 42  ------------------QKELEAKEAARRTAQLYAEYKRQD---------AFQKESIAH 74

Query: 121 QGMEPSRELDSGSKSSEFRTTENVPDIDLKADKTDVATAVPNQETEEIFLVRATDIPTEG 180
                             +T ++      +A K  V T+   + T          + +E
Sbjct: 75  NN----------------KTAKH-----FQAIKGAVMTSEALKPT----------LLSEK 103
```

```
Query: 181 ENVKLGEISELEPVAKEPIRVEDLSKEEEGIALSAKNKHNKRER---RQKADNVAKRIAR 237
           EN  L   ++    A E   +++ + +E  + L+ +  H+ R +   RQ+ +  AK+I+
Sbjct: 104 ENSSLKTTNKRVVQANE---LQETASKESQVPLTIEKGHSVRRKLSKRQQTERAAKKIST 160

Query: 238 ILISIIILVLLLTAFVGYRFVDSAIKPVDSNSNKFVQVEIPIGSGNKLIGQILEKAGVIK 297
           +LIS II+ LL     G  +V SA+ PVD NS+ FVQVEIP GSGNKLIGQIL+K G+IK
Sbjct: 161 VLISSIIITLLAVTLAGAGYVYSALNPVDKNSDAFVQVEIPSGSGNKLIGQILQKKGLIK 220

Query: 298 SATVFNYYSKFKNYSNFQSGYYNLKKSMTLDQIAAELEKGGTAEPTKPALGKILITEGYT 357
           ++TVF++Y+KFKN++NFQSGYYNL+KSM+L++IA+ L++GGTAEPTKP+LGKILI EGYT
Sbjct: 221 NSTVFSFYTKFKNFTNFQSGYYNLQKSMSLEEIASALQEGGTAEPTKPSLGKILIPEGYT 280

Query: 358 IKQIAKAIESN-KIDTKTTSTPYKADDFLKLVQDETFIKKMVAKYPNLLGSLPDKSKAIY 416
           IKQIAKA+E N K  TK   TP+   DFL LV DE FI+ MV +YP LL ++P K KAIY
Sbjct: 281 IKQIAKAVEHNSKGKTKKAKTPFNEKDFLDLVTDEAFIQDMVKRYPKLLATIPTKEKAIY 340

Query: 417 QLEGYLFPATYNYYKDTTLEGLVEDMISTMNTKMAPYYNTIKAKNMSVNDVLTLSSLVEK 476
           +LEGYLFPATYNYYK+TT+  LVEDM++ M+  + PYY+ I A   +VN+VLTL+SLVEK
Sbjct: 341 RLEGYLFPATYNYYKETTMRELVEDMLAAMDATLVPYYDKIAASGKTVNEVLTLASLVEK 400

Query: 477 EGSTDEDRRKIASVFYNRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNI 536
           EGSTD+DRR+IASVFYNRL++G ALQSNIAILYAMGKLG+KT+LAEDA I+T+I SPYNI
Sbjct: 401 EGSTDDDRRQIASVFYNRLNSGMALQSNIAILYAMGKLGEKTTLAEDATIDTTINSPYNI 460

Query: 537 YTNTGLMPGPVDSPSISAIEATIKPASTDYLYFVADVKTGNVYYAKDFETHKANVEKYIN 596
           YTNTGLMPGPV S  +SAIEAT+ PASTDYLYFVA+V TG VYYAK FE H ANVEKY+N
Sbjct: 461 YTNTGLMPGPVASSGVSAIEATLNPASTDYLYFVANVHTGEVYYAKTFEEHSANVEKYVN 520

Query: 597 SQI 599
           SQI
Sbjct: 521 SQI 523
```

A related GBS gene <SEQ ID 8843> and protein <SEQ ID 8844> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -17.88
GvH: Signal Score (-7.5): -3.51
     Possible site: 58
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -13.64 threshold:  0.0
   INTEGRAL    Likelihood =-13.64   Transmembrane  238 - 254 ( 230 - 260)
   PERIPHERAL  Likelihood =  5.78     285
 modified ALOM score:    3.23

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00931(1417 - 2046 of 2400)
GP|8574530|gb|AAF77615.1|AF151720_1|AF151720(5 - 214 of 214) aminodeoxychorismate lyase-
like protein {Streptococcus thermophilus}
%Match = 17.5
%Identity = 64.3  %Similarity = 81.4
Matches = 135  Mismatches = 39  Conservative Sub.s = 36

1236      1266      1296      1326      1356      1386      1416      1446
NYYSKFKNYSNFQSGYYNLKKSMTLDQIAAELEKGGTAEPTKPALGKILITEGYTIKQIAKAIESNKIDTKTTSTPYKAD
                                                                      |.:||  |
                                                                  AKKGKHSSTGLKEK
                                                                          10
```

```
1476      1506      1536      1566      1596      1626      1656      1686
DFLKLVQDETFIKKMVAKYPNLLGSLPDKSKAIYQLEGYLFPATYNYYKDTTLEGLVEDMISTMNTKMAPYYNTIKAKNM
|||||::|: || || |||| || :||: : | | ||||||||||| : |||:| | |:|: ||:| ::||| || : |
DFLKLMKDDAFITKMKAKYPTLLANLPNSTDAKYVLEGYLFPATYNIHDDTTVESLAEEMLFTMDTHLSPYYATILSSNH
             30        40        50        60        70        80        90

1716      1746      1776      1806      1836      1866      1896      1926
SVNDVLTLSSLVEKEGSTDEDRRKIASVFYNRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNIYTNTGL
:||::|||:|||||||:||:||: |||||||||::  |||||||:|| :|||| :|:| ||| |:|:| |||| | : ||
NVNEILTLASLVEKEGATDDDRKNIASVFYNRLNSDMALQSNIAVLYVLGKLGQETTLKEDATIDTNIDSPYNDYVHKGL
             110       120       130       140       150       160       170

1956      1986      2016      2046      2076      2106      2136      2166
MPGPVDSPSISAIEATIKPASTDYLYFVADVKTGNVYYAKDFETHKANVEKYINSQIN*AYKHGASHHVYIFDLKK*KEK
|||||||||:||||| | |:|| |:|||||| |||||:|:
MPGPVDSPSLSAIEAVINPSSTKYMYFVADVSTGNVYFAE
             190       200       210
```

SEQ ID 8844 (GBS370) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 6; MW 70kDa).

GBS370-His was purified as shown in Figure 209, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1557

A DNA sequence (GBSx1649) was identified in *S.agalactiae* <SEQ ID 4807> which encodes the amino acid sequence <SEQ ID 4808>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0183(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10077> which encodes amino acid sequence <SEQ ID 10078> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA98889 GB:Z74367 ORF YDR071c [Saccharomyces cerevisiae]
 Identities = 52/174 (29%), Positives = 81/174 (45%), Gaps = 18/174 (10%)

Query: 27  MSMIIRNGCLEDLQQVISIEQINFSEAEAASKKAMQERLTIMTDT---FLVAEINGR--- 80
           + M IR   +EDL+Q++++E   F   E AS++ +  RL    +      + EI G+
Sbjct: 10  LHMYIRPLIIEDLKQILNLESQGFPPNERASEEIISFRLINCPELCSGLFIREIEGKEVK 69

Query: 81  ---LAGYIEGPVIKGRYLTDDLFHKVSEFPVRVGGFIGITSLSIHPDFKGQGIGTALLAA 137
              L G+I G  I   Y+T +   K+    V     IGI S+ I P+++ + + T LL
Sbjct: 70  KETLIGHIMGTKIPHEYITIESMGKLQ---VESSNHIGIHSVVIKPEYQKKNLATLLLTD 126

Query: 138 MKDLVVSQE-RDGISLTCHDDLISFYEMNGFKDEGES-----DSKHGGSLWYNM 185
               + +QE  + I L  H+ LI FYE  GFK   E+     D       W +M
Sbjct: 127 YIQKLSNQEIGNKIVLIAHEPLIPFYERVGFKIIAENTNVAKDKNFAEQKWIDM 180
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4809> which encodes the amino acid sequence <SEQ ID 4810>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 87/159 (54%), Positives = 117/159 (72%), Gaps = 1/159 (0%)

Query: 29  MIIRNGCLEDLQQVISIEQINFSEAEAASKKAMQERLTIMTDTFLVAEINGRLAGYIEGP 88
           M+IR     DL+ + +IE  NFS  EA ++  ++E + ++ DTFLVA I+  + GYIEGP
Sbjct: 1   MLIRQVQGSDLEVIATIESDNFSPQEATTRAVLEEHIRLIPDTFLVALIDQEIVGYIEGP 60

Query: 89  VIKGRYLTDDLFHKVSEFPVRVGGFIGITSLSIHPDFKGQGIGTALLAAMKDLVVSQERD 148
           V+    L D LFH V++ P + GG+I ITSLSI   F+ QG+GTALLAA+KDLVV+Q+R
Sbjct: 61  VVTTPILEDSLFHGVTKNP-KTGGYIAITSLSIAKHFQQQGVGTALLAALKDLVVAQQRT 119

Query: 149 GISLTCHDDLISFYEMNGFKDEGESDSKHGGSLWYNMIW 187
           G+ LTCHD LIS+YEMNGF ++G S+S+HGG+LWY MIW
Sbjct: 120 GLILTCHDYLISYYEMNGFINQGISESQHGGTLWYQMIW 158
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1558

A DNA sequence (GBSx1650) was identified in *S.agalactiae* <SEQ ID 4811> which encodes the amino acid sequence <SEQ ID 4812>. This protein is predicted to be udp-n-acetylmuramate--alanine ligase (murC/ddlA). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -2.60   Transmembrane  272 - 288 ( 270 - 288)

----- Final Results -----
               bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00294 GB:AF008220 putative UDP-N-acetylmuramate-alanine
          ligase [Bacillus subtilis]
 Identities = 238/432 (55%), Positives = 315/432 (72%), Gaps = 3/432 (0%)

Query: 5   YHFIGIKGSGMSALALMLHQMGHNVQGSDVDKYYFTQRGLEQAGVTILPFSPNNISEDLE 64
           YHF+GIKG+GMS LA +LH  G+ VQGSD++K+ FTQ  LE+  +TILPFS  NI   +
Sbjct: 4   YHFVGIKGTGMSPLAQILHDNGYTVQGSDIEKFIFTQTALEKRNITILPFSAENIKPGMT 63

Query: 65  IIAGNAFRPDNNEELAYVIEKGYQFKRYHEFLGDFMRQFTSLGVAGAHGKTSTTGLLAHV 124
           +IAGNAF PD + E+   + +G    RYH+FLGD+M++FTS+ V GAHGKTSTTGLLAHV
Sbjct: 64  VIAGNAF-PDTHPEIEKAMSEGIPVIRYHKFLGDYMKKFTSVAVTGAHGKTSTTGLLAHV 122

Query: 125 LKNITDTSFLIGDGTGRGSANANYFVFEADEYERHFMPYHPEYSIITNIDFDHPDYFTGL 184
           ++N   TSFLIGDGTG+G+ N+ YFVFEA EY RHF+ Y P+Y+I+TNIDFDHPDYF+ +
Sbjct: 123 IQNAKPTSFLIGDGTGQGNENSEYFVFEACEYRRHFLSYQPDYAIMTNIDFDHPDYFSSI 182

Query: 185 EDVFNAFNDYAKQVQKGLFIYGEDPKLHEITSEAPIYYYGFEDSNDFIAKDITRTVNGSD 244
           +DVF+AF + A QV KG+   G+D  L +I +  P+ YYG  + NDF A++I ++  G+
Sbjct: 183 DDVFDAFQEMALQVNKGIIACGDDEHLPKIHANVPVVYYGTGEENDFQARNIVKSTEGTT 242

Query: 245 FKVFYNQEEIGQFHVPAYGKHNILNATAVIANLYIMGIDMALVAEHLKTFSGVKRRFTEK 304
           F VF        F++PAYG HN+LN+ AVIA  +   ID +++   LK+F GVKRRF EK
Sbjct: 243 FDVFVRNTFYDTFYIPAYGHHNVLNSLAVIALCHYEEIDSSIIKHALKSFGGVKRRFNEK 302
```

```
Query: 305 IIDDTVIIDDFAHHPTEIIATLDAARQKYPSKEIVAIFQPHTFTRTIALLDEFAHALSQA 364
            + D V+IDD+AHHPTEI  T++AARQKYP +EIVA+FQPHTFTRT   LDEFA +LS A
Sbjct: 303 QLGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHTFTRTQQFLDEFAESLSGA 362

Query: 365 DSVYLAQIYGSAREVDNGEVKVEDLAAKIVKHSDLVTVENVSPLLNHDNAVYVFMGAGDI 424
           D VYL  I+GSARE + G++ + DL  KI  ++ L+  ++ S L  HD AV +FMGAGDI
Sbjct: 363 DCVYLCDIFGSARE-NAGKLTIGDLQGKI-HNAKLIEEDDTSVLKAHDKAVLIFMGAGDI 420

Query: 425 QLYERSFEELLA 436
           Q Y R++E ++A
Sbjct: 421 QKYMRAYENVMA 432
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4813> which encodes the amino acid sequence <SEQ ID 4814>. Analysis of this protein sequence reveals the following:

```
      Possible site: 31
  >>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood = -4.57   Transmembrane  271 - 287 ( 269 - 288)

  ----- Final Results -----
               bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC00294 GB:AF008220 putative UDP-N-acetylmuramate-alanine
           ligase [Bacillus subtilis]
 Identities = 236/431 (54%), Positives = 310/431 (71%), Gaps = 2/431 (0%)

Query: 5   YHFIGIKGSGMSALALMLHQMGHKVQGSDVEKYYFTQRGLEQAGITILPFSEDNITPDME 64
           YHF+GIKG+GMS LA +LH  G+ VQGSD+EK+ FTQ  LE+  ITILPFS +NI P M
Sbjct: 4   YHFVGIKGTGMSPLAQILHDNGYTVQGSDIEKFIFTQTALEKRNITILPFSAENIKPGMT 63

Query: 65  LIVGNAFRENNKEVAYALRHQIPFKRYHDFLGDFMKSFISFAVAGAHGKTSTTGLLSHVL 124
           +I GNAF + + E+  A+   IP  RYH FLGD+MK F S AV GAHGKTSTTGLL+HV+
Sbjct: 64  VIAGNAFPDTHPEIEKAMSEGIPVIRYHKFLGDYMKKFTSVAVTGAHGKTSTTGLLAHVI 123

Query: 125 KNITDTSYLIGDGTGRGSANAQYFVFESDEYERHFMPYHPEYSIITNIDFDHPDYFTGIA 184
           +N   TS+LIGDGTG+G+ N++YFVFE+ EY RHF+ Y P+Y+I+TNIDFDHPDYF+ I
Sbjct: 124 QNAKPTSFLIGDGTGQGNENSEYFVFEACEYRRHFLSYQPDYAIMTNIDFDHPDYFSSID 183

Query: 185 DVRNAFNDYAKQVKKALFVYGEDDELKKIEAPAPIYYYGFEEGNDFIAYDITRTTNGSDF 244
           DV +AF + A QV K +   G+D+ L KI A  P+ YYG  E NDF A +I ++T G+ F
Sbjct: 184 DVFDAFQEMALQVNKGIIACGDDEHLPKIHANVPVVYYGTGEENDFQARNIVKSTEGTTF 243

Query: 245 KVKHQGEVIGQFHVPAYGKHNILNATAVIANLFVAGIDMALVADHLKTFSGVKRRFTEKI 304
            V  +      F++PAYG HN+LN+ AVIA      ID +++   LK+F GVKRRF EK
Sbjct: 244 DVFVRNTFYDTFYIPAYGHHNVLNSLAVIALCHYEEIDSSIIKHALKSFGGVKRRFNEKQ 303

Query: 305 INDTIIIDDFAHHPTEIVATIDAARQKYPSKEIVAIFQPHTFTRTIALLEDFACALNEAD 364
           + D ++IDD+AHHPTEI  TI+AARQKYP +EIVA+FQPHTFTRT   L++FA +L+ AD
Sbjct: 304 LGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHTFTRTQQFLDEFAESLSGAD 363

Query: 365 SVYLAQIYGSAREVDKGEVKVEDLAAKIIKPSQVVTVENVSPLLDHDNAVYVFMGAGDIQ 424
            VYL  I+GSARE + G++ + DL  K I  ++++  ++ S L  HD AV +FMGAGDIQ
Sbjct: 364 CVYLCDIFGSARE-NAGKLTIGDLQGK-IHNAKLIEEDDTSVLKAHDKAVLIFMGAGDIQ 421

Query: 425 LYEHSFEELLA 435
            Y  ++E ++A
Sbjct: 422 KYMRAYENVMA 432
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 369/443 (83%), Positives = 406/443 (91%), Gaps = 1/443 (0%)
```

```
Query: 1   MSKTYHFIGIKGSGMSALALMLHQMGHNVQGSDVDKYYFTQRGLEQAGVTILPFSPNNIS 60
           MSKTYHFIGIKGSGMSALALMLHQMGH VQGSDV+KYYFTQRGLEQAG+TILPFS +NI+
Sbjct: 1   MSKTYHFIGIKGSGMSALALMLHQMGHKVQGSDVEKYYFTQRGLEQAGITILPFSEDNIT 60

Query: 61  EDLEIIAGNAFRPDNNEELAYVIEKGYQFKRYHEFLGDFMRQFTSLGVAGAHGKTSTTGL 120
            D+E+I GNAFR +NN+E+AY +      FKRYH+FLGDFM+ F S  VAGAHGKTSTTGL
Sbjct: 61  PDMELIVGNAFR-ENNKEVAYALRHQIPFKRYHDFLGDFMKSFISFAVAGAHGKTSTTGL 119

Query: 121 LAHVLKNITDTSFLIGDGTGRGSANANYFVFEADEYERHFMPYHPEYSIITNIDFDHPDY 180
           L+HVLKNITDTS+LIGDGTGRGSANA YFVFE+DEYERHFMPYHPEYSIITNIDFDHPDY
Sbjct: 120 LSHVLKNITDTSYLIGDGTGRGSANAQYFVFESDEYERHFMPYHPEYSIITNIDFDHPDY 179

Query: 181 FTGLEDVFNAFNDYAKQVQKGLFIYGEDPKLHEITSEAPIYYYGFEDSNDFIAKDITRTV 240
           FTG+ DV NAFNDYAKQV+K LF+YGED +L +I + APIYYYGFE+ NDFIA DITRT
Sbjct: 180 FTGIADVRNAFNDYAKQVKKALFVYGEDDELKKIEAPAPIYYYGFEEGNDFIAYDITRTT 239

Query: 241 NGSDFKVFYNQEEIGQFHVPAYGKHNILNATAVIANLYIMGIDMALVAEHLKTFSGVKRR 300
           NGSDFKV +  E IGQFHVPAYGKHNILNATAVIANL++ GIDMALVA+HLKTFSGVKRR
Sbjct: 240 NGSDFKVKHQGEVIGQFHVPAYGKHNILNATAVIANLFVAGIDMALVADHLKTFSGVKRR 299

Query: 301 FTEKIIDDTVIIDDFAHHPTEIIATLDAARQKYPSKEIVAIFQPHTFTRTIALLDEFAHA 360
           FTEKII+DT+IIDDFAHHPTEI+AT+DAARQKYPSKEIVAIFQPHTFTRTIALL++FA A
Sbjct: 300 FTEKIINDTIIIDDFAHHPTEIVATIDAARQKYPSKEIVAIFQPHTFTRTIALLEDFACA 359

Query: 361 LSQADSVYLAQIYGSAREVDNGEVKVEDLAAKIVKHSDLVTVENVSPLLNHDNAVYVFMG 420
           L++ADSVYLAQIYGSAREVD GEVKVEDLAAKI+K S +VTVENVSPLL+HDNAVYVFMG
Sbjct: 360 LNEADSVYLAQIYGSAREVDKGEVKVEDLAAKIIKPSQVVTVENVSPLLDHDNAVYVFMG 419

Query: 421 AGDIQLYERSFEELLANLTKNTQ 443
           AGDIQLYE SFEELLANLTKN Q
Sbjct: 420 AGDIQLYEHSFEELLANLTKNNQ 442
```

SEQ ID 4812 (GBS157) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 11; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 8; MW 74kDa), Figure 33 (lane 8; MW 74kDa) and Figure 37 (lane 3; MW 74kDa).

The GBS157-GST fusion product was purified (Figure 112A; see also Figure 200, lane 3) and used to immunise mice (lane 1+2 product; 19.5μg/mouse). The resulting antiserum was used for Western blot (Figure 112B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 4812 (GBS157) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 11-13; MW 74kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1559

A DNA sequence (GBSx1651) was identified in *S.agalactiae* <SEQ ID 4815> which encodes the amino acid sequence <SEQ ID 4816>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1980(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4817> which encodes the amino acid sequence <SEQ ID 4818>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2731(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 80/201 (39%), Positives = 126/201 (61%), Gaps = 9/201 (4%)

Query: 7   RFPLIADDEPVMSPLVKMNLYDNEDLINNIRDFYQEKTYQSMVKSNYEHEEISHPKVIEN 66
           +FPL+AD   +  P  +M LY+NEDLI NIR +YQ+K Y  + ++    EE +
Sbjct: 5   QFPLVADGIAISDPAKQMALYENEDLITNIRGYYQDKEYDDIARN----EEFTAKATSRQ 60

Query: 67  DPVPPQ--SFVKKATELSKSRQEAKRSVREKRQAYYAKQEFKAPSKEAFQQQLKATVPKK 124
            P   +  S  +K   + ++RQ+AK+ ++EKRQAY AK+    P + + +QQ   + P +
Sbjct: 61  TPSSKRFCSNDEKHHYVKEARQKAKQDLKEKRQAYLAKEMAYVPKQVSKKQQPADSSPSQ 120

Query: 125 QTQRKVTELSHLSDRLQQESYILAEIPIIFQEPDNTPNP-KTKKNNFDFLKRSQVYNKQD 183
           +  +  TE+S  + +L Q++YILAE+P  ++EP N P    TKKNN+DFLK SQ+YN ++
Sbjct: 121 K--QATTEMSRFTKKLHQDNYILAELPKEYKEPKNLPQQGTTKKNNYDFLKSSQIYNNKE 178

Query: 184 NQFHKERAKAQELNLTRFKDI 204
            +  +E+  AQELNL+RF+D+
Sbjct: 179 MRQQREKTIAQELNLSRFEDL 199
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1560

A DNA sequence (GBSx1652) was identified in *S.agalactiae* <SEQ ID 4819> which encodes the amino acid sequence <SEQ ID 4820>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4959(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1561

A DNA sequence (GBSx1653) was identified in *S.agalactiae* <SEQ ID 4821> which encodes the amino acid sequence <SEQ ID 4822>. This protein is predicted to be SNF. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane  743 - 759 ( 743 - 759)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA67095 GB:X98455 SNF [Bacillus cereus]
 Identities = 259/678 (38%), Positives = 406/678 (59%), Gaps = 21/678 (3%)

Query: 369  QNEILLQMVFDYGNDLTVHNRQELEQLTFASHFKHEEKVFKLLEKYGFAPHFSTSHPAYS 428
            +N +L  + F YGN +     ++ +   F    K E+++  ++ +  FA      +  ++
Sbjct: 388  KNRLLAGLEFHYGNVVINPLEEDGQPSVFNRDEKKEKEILDIMSESAFAKT-EGGYFMHN 446

Query: 429  AQELYDFYTYMLPQFKKMGTV--SLSAKLESYRLIERPQIDIEAKGSL--LDISFDFSDL 484
             +  Y+F  +++P  K +  +  + + KL  ++    P I +  K  +  L   FD   +
Sbjct: 447  EEAEYNFLYHIVPTLKGLVDIYATTAIKLRIHKGDTAPLIRVRRKERIDWLSFRFDIKGI 506

Query: 485  LENDVDQALVALFDNNPYFVNKSGQLVIFD-EETKKVSATLQ--GLRARRAKNGHIELDN 541
             E ++   L AL +   Y+   +G L+  + +E  +++  ++  G+R        + +
Sbjct: 507  PEAEIKGVLAALEEKRKYYRLANGSLLSLESKEFNEINQFVKESGIRKEFLHGEEVNVPL 566

Query: 542  IAAFQLSELFANQDNVSFSQHFYQLIEDLRHPEKFK--IPGLSVSASLRDYQLTGVRWLS 599
            I + +        + +S  +    L+E +++P+K K  +P  ++ A +R+YQ+ G  W+
Sbjct: 567  IRSVKWMNGLHEGNVLSLDESVQDLVESIQNPKKLKFTVPP-TLHAVMREYQVYGFEWMK 625

Query: 600  MLDHYGFAGILADDMGLGKTLQTISFLSTKLT--RDSR--VLILSPSSLIYNWQDEFHKF 655
             L +Y F GILADDMGLGKTLQ+I+++ + L   R+ +  +L++SPSSL+YNW  E  KF
Sbjct: 626  TLAYYRFGGILADDMGLGKTLQSIAYIDSVLPEIREKKLPILVVSPSSLVYNWFSELKKF 685

Query: 656  APDVDVAVAYGSKIRRDEIIAE--RHQVIITSYSSFRQDFETYSEGNYDYLILDEAQVMK 713
            AP +   +A G++  R +I+ +     V+ITSY   R+D  +Y+   +  L LDEAQ  K
Sbjct: 686  APHIRAVIADGNQTERRKILKDVAEFDVVITSYPLLRRDVRSYARP-FHTLFLDEAQAFK 744

Query: 714  NAQTKIAHSLRSFEVKNCFALSGTPIENKLLEIWSIFQIILPGLLPGKKEFLKLNPKQVA 773
            N  T+ A ++++ + +  F L+GTP+EN L E+WSIF ++ P LLPG+KEF  L  + +A
Sbjct: 745  NPTTQTARAVKTIQAEYRFGLTGTPVENSLEELWSIFHVVFPELLPGRKEFGDLRREDIA 804

Query: 774  RYIKPFVMRRRKEEVLPELPDLIEMNYPNEMTDSQKVIYLAQLRQI-QESIQHSSDADLN 832
              +KPFV+RR KE+VL ELPD IE    +E+   QK +Y A L ++ +E+++H     L
Sbjct: 805  NAVKPFVLRRLKEDVLQELPDKIEHLQSSELLPDQKRLYAAYLAKLREETLKHLDKDTLR 864

Query: 833  RRKIEILSGITRLRQICDTPRLFMD-YDGESGKLESLRQLLTQIKENGHRALIFSQFRGM 891
            + KI IL+G+TRLRQIC+ P LF+D Y G S KLE L  +L + +  G R LIFSQF  M
Sbjct: 865  KNKIRILAGLTRLRQICNHPALFVDDYKGSSAKLEQLLDILEECRSTGKRILIFSQFTKM 924

Query: 892  LDIAEREMVAMGLTTYKITGSTPANERHEMTRAFNAGSKDAFLISLKAGGVGLNLTGADT 951
            L I  RE+    +  + + G+TP+ ER E+   FN G  D FLISLKAGG GLNLTGADT
Sbjct: 925  LSIIGRELNRQAIPYFYLDGNTPSQERVELCNRFNEGEGDLFLISLKAGGTGLNLTGADT 984

Query: 952  VVLIDLWWNPAVEMQAISRAHRLGQKENVEVYRLITRGTIEEKILEMQETKKHLVTTVLD 1011
            V+L DLWWNPAVE QA  RA+R+GQK  V+V +L+  GTIEEK+ E+QE+KKHL+  V++
Sbjct: 985  VILYDLWWNPAVEQQAADRAYRMGQKNTVQVIKLVAHGTIEEKMHELQESKKHLIAEVIE 1044

Query: 1012 -GNETHASMSVDDIREIL 1028
             G E  +S++ ++IR+IL
Sbjct: 1045 PGEEKLSSITEEEIRDIL 1062
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4823> which encodes the amino acid sequence <SEQ ID 4824>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3909(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 674/1031 (65%), Positives = 834/1031 (80%), Gaps = 2/1031 (0%)

Query: 1    MSRMIPGRIRNQGIELYEQGLVSLISQEGNLLKAKVGDCQIEYSLVTEETKCSCDFFARK 60
            M+R+IPGR+RN+GI+LYEQGLVS       +L+ +V   Q++Y    E+  C CD F  K
Sbjct: 2    MARLIPGRVRNEGIKLYEQGLVSFQDDNKGILQIEVETYQVQYGADDEDITCQCDTFHMK 61

Query: 61   GYCQHLAALEHFLKNDPEGKAILSKVQVQQESQQETKKKTSFGSVFLDSLIINEDDTIKY 120
             YC+H+AA+E+FLKND +GK  L ++  Q + ++ TKK TSFGS+FLDSL +NEDD++KY
Sbjct: 62   HYCKHIAAVEYFLKNDQKGKLFLKQLTNQTKIKETTKKMTSFGSLFLDSLAMNEDDSVKY 121

Query: 121  QLSAQGEQNPYANDIWWTLKIRRLPDDRSYVIRDIKAFLNTVRKEAYYQIGKQYFETLSL 180
            +LSA G ++P+++D WW+LKI RLPDDRSYVIRDIK FL  ++KE +YQIGK YFE LS
Sbjct: 122  RLSALGSRSPFSSDYWWSLKINRLPDDRSYVIRDIKGFLQLIKKEGFYQIGKNYFEQLSW 181

Query: 181  IQFDETSQELIEFLWRLIPSHSSKIDLEFILPNQGRHLSLTRGFFEEGVTLMNALENFSF 240
            +QFD +SQ LIEFLWRL  S + K D E I PN  RHL L  GFFEEG+  + +L +F+F
Sbjct: 182  LQFDPSSQALIEFLWRLA-SDTDKGDNENIFPNHARHLRLPSGFFEEGIHYLTSLYDFTF 240

Query: 241  ESDFHQFNHLYFKELEGEDHLYQFKVIVHRQSIELEIKEKDLKPLFANSYLFYRDTFYHL 300
            E     ++HL+ + LE E  LY+FKV VHR+SIEL+I EK+++ LF N YL Y+DTFYHL
Sbjct: 241  EGPSQTYHHLFVRSLEAEAGLYEFKVEVHRKSIELQIAEKNVQYLFDNDYLLYQDTFYHL 300

Query: 301  NLKQEKMVTAIRSLPIEGDLAKHIHFDLDDQDKLAAHLLDFKEIGLVDAPRSFSIHDFKV 360
             LKQ KMV AIRSLPIE DLAKHIHFDLDD  KLAA L DFK+IGLV+AP+SF+I DF+V
Sbjct: 301  TLKQRKMVQAIRSLPIEADLAKHIHFDLDDHAKLAASLSDFKQIGLVEAPKSFAIRDFEV 360

Query: 361  NFEFDINSQNEILLQMVFDYGNDLTVHNRQELEQLTFASHFKHEEKVFKLLEKYGFAPHF 420
             F+FD+ +++EI  Q++FDYGN   V ++  LE L FASH K EEK+ + L  +GF+P F
Sbjct: 361  TFQFDLLNRDEISCQLMFDYGN-YQVSDKASLEALPFASHLKKEEKINRSLLAFGFSPQF 419

Query: 421  STSHPAYSAQELYDFYTYMLPQFKKMGTVSLSAKLESYRLIERPQIDIEAKGSLLDISFD 480
             +     SA+ELY F+   +P F+++G V+LS  +++ ++ E P+I I     LLDISFD
Sbjct: 420  YSKKRLTSAKELYTFFEETVPCFERLGNVALSTAIQALQVKEMPKIAIRRNQGLLDISFD 479

Query: 481  FSDLLENDVDQALVALFDNNPYFVNKSGQLVIFDEETKKVSATLQGLRARRAKNGHIELD 540
            FS ++END+DQA+ ALF NNPYFV+++GQLV+FD+ET+KVS +LQ LRAR+ KNGH++LD
Sbjct: 480  FSTIIENDIDQAVTALFQNNPYFVSQTGQLVVFDDETQKVSKSLQELRARQLKNGHLQLD 539

Query: 541  NIAAFQLSELFANQDNVSFSQHFYQLIEDLRHPEKFKIPGLSVSASLRDYQLTGVRWLSM 600
             I A Q+S+LF    +V FS+   +L   L+HPE F I  L V A +RDYQ  GV+WLSM
Sbjct: 540  GIRALQVSKLFEGMTSVHFSKELEELAYHLQHPETFSIKPLPVKAQMRDYQRNGVQWLSM 599

Query: 601  LDHYGFAGILADDMGLGKTLQTISFLSTKLTRDSRVLILSPSSLIYNWQDEFHKFAPDVD 660
            L+HYGF GILADDMGLGKTLQT++FL++ L  DS+VLILSPSSLIYNW DE  KF P +D
Sbjct: 600  LNHYGFGGILADDMGLGKTLQTLAFLASHLKSDSKVLILSPSSLIYNWFDECQKFTPQLD 659

Query: 661  VAVAYGSKIRRDEIIAERHQVIITSYSSFRQDFETYSEGNYDYLILDEAQVMKNAQTKIA 720
            V V+YG K  RD+II E HQ+ ITSYSSFRQDFETY   +YDYLILDEAQV+KNAQTKI+
Sbjct: 660  VVVSYGLKQIRDQIIEEGHQITITSYSSFRQDFETYQAFHYDYLILDEAQVIKNAQTKIS 719

Query: 721  HSLRSFEVKNCFALSGTPIENKLLEIWSIFQIILPGLLPGKKEFLKLNPKQVARYIKPFV 780
            H LR+F   NCFALSGTPIENK+LEIWSIFQI+LPGLLP KKEFLKL  +QV+RYIKPFV
Sbjct: 720  HCLRAFNTANCFALSGTPIENKMLEIWSIFQIVLPGLLPTKKEFLKLTAEQVSRYIKPFV 779

Query: 781  MRRRKEEVLPELPDLIEMNYPNEMTDSQKVIYLAQLRQIQESIQHSSDADLNRRKIEILS 840
            MRR+KE+VLPELPDLIE+NY NEMTD QK IYLAQLRQ+Q+ I++SSD D++R+KIEILS
Sbjct: 780  MRRKKEDVLPELPDLIEINYSNEMTDEQKAIYLAQLRQMQDQIRNSSDVDISRQKIEILS 839

Query: 841  GITRLRQICDTPRLFMDYDGESGKLESLRQLLTQIKENGHRALIFSQFRGMLDIAEREMV 900
            GITRLRQICDTP LFMDY G+SGKL+SLR LLTQIKENGHRALIFSQFRGMLD+A++EM
```

```
Sbjct: 840   GITRLRQICDTPSLFMDYQGKSGKLDSLRILLTQIKENGHRALIFSQFRGMLDLAKQEMT 899

Query: 901   AMGLTTYKITGSTPANERHEMTRAFNAGSKDAFLISLKAGGVGLNLTGADTVVLIDLWWN 960
              A+GLT+Y++TGSTPANER EMTRAFN GSKDAFLISLKAGGVG+NLTGADTV+LIDLWWN
Sbjct: 900   ALGLTSYQMTGSTPANERQEMTRAFNNGSKDAFLISLKAGGVGINLTGADTVILIDLWWN 959

Query: 961   PAVEMQAISRAHRLGQKENVEVYRLITRGTIEEKILEMQETKKHLVTTVLDGNETHASMS 1020
              PAVEMQAISRA+R+GQKENVEVYRLITRGTIEEKILE+QE+K++LVTTVLDGNE+ ASMS
Sbjct: 960   PAVEMQAISRAYRIGQKENVEVYRLITRGTIEEKILELQESKRNLVTTVLDGNESRASMS 1019

Query: 1021 VDDIREILGVS 1031
             +++I+EILG++
Sbjct: 1020 IEEIKEILGLN 1030
```

SEQ ID 4822 (GBS369) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 5; MW 120kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 6; MW 142kDa).

The GBS369-GST fusion product was purified (Figure 215, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 303), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1562

A DNA sequence (GBSx1654) was identified in *S.agalactiae* <SEQ ID 4825> which encodes the amino acid sequence <SEQ ID 4826>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 1034:

```
 Identities = 34/38 (89%), Positives = 37/38 (96%)

Query: 1  MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRTACQFS 38
           +EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRTAC+ S
Sbjct: 1  LEKEAKQMIDLKRNLFKIDVRAQKDEEKVFMRTACRQS 38
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1563

A DNA sequence (GBSx1656) was identified in *S.agalactiae* <SEQ ID 4827> which encodes the amino acid sequence <SEQ ID 4828>. This protein is predicted to be phosphoglycerate dehydrogenase (era2). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3709(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA88823 GB:AB016077 phosphoglycerate dehydrogenase
           [Streptococcus mutans]
 Identities = 377/436 (86%), Positives = 414/436 (94%)

Query: 1   MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYTTGEWLNRKFSLIDTGG 60
           M LPTVAIVGRPNVGKS LFNRIAGERISIVEDVEGVTRDRIYT  EWLNR+FS+IDTGG
Sbjct: 1   MALPTVAIVGRPNVGKSALFNRIAGERISIVEDVEGVTRDRIYTKAEWLNRQFSIIDTGG 60

Query: 61  IDDVDAPFMEQIKHQADIAMTEADVIVFVVSGKEGVTDADEYVSRILYKTNKPVILAVNK 120
           IDDVDAPFMEQIKHQADIAMTEADVIVFVVS KEG+TDADEYV++ILY+T+KPVILAVNK
Sbjct: 61  IDDVDAPFMEQIKHQADIAMTEADVIVFVVSAKEGITDADEYVAKILYRTHKPVILAVNK 120

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILDAIVENLPVEEENENPDIIRFSLI 180
           VDNPEMR+ IYDFY+LGLGDPYP+SS HGIGTGD+LDAIV+NLP E + E+ DII+FSLI
Sbjct: 121 VDNPEMRSAIYDFYALGLGDPYPVSSAHGIGTGDVLDAIVDNLPAEAQEESSDIIKFSLI 180

Query: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTNFVDSQGQEYTMIDTAGMRKSGKVY 240
           GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDT F D +GQE+TMIDTAGMRKSGKVY
Sbjct: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTTFTDEEGQEFTMIDTAGMRKSGKVY 240

Query: 241 ENTEKYSVMRSMRAIDRSDVVLMVINAEEGIREYDKRIAGFAHETGKGIIIVVNKWDTIE 300
           ENTEKYSVMR+MRAIDRSD+VLMV+NAEEGIREYDKRIAGFAHE GKGI++VVNKWD I+
Sbjct: 241 ENTEKYSVMRAMRAIDRSDIVLMVLNAEEGIREYDKRIAGFAHEAGKGIVVVVNKWDAIK 300

Query: 301 KDNHTVSQWEADIRDNFQFLSYAPIIFVSAETKQRLHKLPDMIKRISESQNKRIPSAVLN 360
           KDN TV+QWE DIRDNFQ++ YAPI+FVSA TKQRLHKLPD+IK++S+SQN RIPS+VLN
Sbjct: 301 KDNRTVAQWETDIRDNFQYIPYAPIVFVSAVTKQRLHKLPDVIKQVSQSQNTRIPSSVLN 360

Query: 361 DVIMDAIAINPTPTDKGKRLKIFYATQVAVKPPTFVVFVNEEELMHFSYLRFLENQIREA 420
           DV+MDA+AINPTPTDKGKRLKIFYATQV+VKPPTFV+FVNEEELMHFSYLRFLENQIR+A
Sbjct: 361 DVVMDAVAINPTPTDKGKRLKIFYATQVSVKPPTFVIFVNEEELMHFSYLRFLENQIRQA 420

Query: 421 FVFEGTPINLIARKRK 436
           FVFEGTPI LIARKRK
Sbjct: 421 FVFEGTPIRLIARKRK 436
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4829> which encodes the amino acid sequence <SEQ ID 4830>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3463(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 403/436 (92%), Positives = 422/436 (96%)

Query: 1   MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYTTGEWLNRKFSLIDTGG 60
           MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIY TGEWLNR+FSLIDTGG
Sbjct: 1   MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYATGEWLNRQFSLIDTGG 60

Query: 61  IDDVDAPFMEQIKHQADIAMTEADVIVFVVSGKEGVTDADEYVSRILYKTNKPVILAVNK 120
           IDDVDAPFMEQIKHQA IAM EADVIVFVVSGKEGVTDADEYVS+ILY+TN PVILAVNK
Sbjct: 61  IDDVDAPFMEQIKHQAQIAMEEADVIVFVVSGKEGVTDADEYVSKILYRTNTPVILAVNK 120

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILDAIVENLPVEEENENPDIIRFSLI 180
```

```
           VDNPEMRNDIYDFYSLGLGDPYP+SSVHGIGTGD+LDAIVENLPVEE  EN DIIRFSLI
Sbjct: 121 VDNPEMRNDIYDFYSLGLGDPYPVSSVHGIGTGDVLDAIVENLPVEEAEENDDIIRFSLI 180

Query: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTNFVDSQGQEYTMIDTAGMRKSGKVY 240
           GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDT+F D+ GQE+TMIDTAGMRKSGK+Y
Sbjct: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTHFTDADGQEFTMIDTAGMRKSGKIY 240

Query: 241 ENTEKYSVMRSMRAIDRSDVVLMVINAEEGIREYDKRIAGFAHETGKGIIIVVNKWDTIE 300
           ENTEKYSVMR+MRAIDRSDVVLMVINAEEGIREYDKRIAGFAHE GKG+IIVVNKWDTI+
Sbjct: 241 ENTEKYSVMRAMRAIDRSDVVLMVINAEEGIREYDKRIAGFAHEAGKGMIIVVNKWDTID 300

Query: 301 KDNHTVSQWEADIRDNFQFLSYAPIIFVSAETKQRLHKLPDMIKRISESQNKRIPSAVLN 360
           KDNHTV++WEADIRD FQFL+YAPIIFVSA TKQRL+KLPD+IKRISESQNKRIPSAVLN
Sbjct: 301 KDNHTVAKWEADIRDQFQFLTYAPIIFVSALTKQRLNKLPDLIKRISESQNKRIPSAVLN 360

Query: 361 DVIMDAIAINPTPTDKGKRLKIFYATQVAVKPPTFVVFVNEEELMHFSYLRFLENQIREA 420
           DVIMDAIAINPTPTDKGKRLKIFYATQV+VKPPTFVVFVNEEELMHFSYLRFLENQIR A
Sbjct: 361 DVIMDAIAINPTPTDKGKRLKIFYATQVSVKPPTFVVFVNEEELMHFSYLRFLENQIRAA 420

Query: 421 FVFEGTPINLIARKRK 436
           F FEGTPI+LIARKRK
Sbjct: 421 FTFEGTPIHLIARKRK 436
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1564

A DNA sequence (GBSx1657) was identified in *S.agalactiae* <SEQ ID 4831> which encodes the amino acid sequence <SEQ ID 4832>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2734(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00359 GB:AF008220 DnaI [Bacillus subtilis]
 Identities = 105/313 (33%), Positives = 191/313 (60%), Gaps = 17/313 (5%)

Query: 1   MKSVGQALENQGRVP--RNTNDELIQMILADAQVAEFIKTHQ--LSQREINISMSKFNQF 56
           M+ +G++L+     P  +   +++ + ++ D  V  F+K ++  + Q+ I  S++K  ++
Sbjct: 1   MEPIGRSLQGVTGRPDFQKRLEQMKEKVMKDQDVQAFLKENEEVIDQKMIEKSLNKLYEY 60

Query: 57  LIERQK-----FKNKDSQYIAKGYEPILVMNEGYADVSYLE--TRELIEAQKKQAISDRI 109
            IE+ K      ++++   + +GY P LV+N    D+ Y E   +  ++ QKKQ     +
Sbjct: 61  -IEQSKNCSYCSEDENCNNLLEGYHPKLVVNGRSIDIEYYECPVKRKLDQQKKQ--QSLM 117

Query: 110 NLVNLPKSYRNIRMTDFDINNESRMKAMSQLLDFVETYPSYNH-KGLYLYGDMGVGKSYL 168
             + + +          DI++ SR+     + DF+++Y      KGLYLYG  GVGK+++
Sbjct: 118 KSMYIQQDLLGATFQQVDISDPSRLAMFQHVTDFLKSYNETGKGKGLYLYGKFGVGKTFM 177

Query: 169 MAAMARELSERKGVSTTLLHFPSFAIDVKNAISSGTVKDEIDAVKSVPILILDDIGAEQA 228
           +AA+A EL+E++  S+ +++ P F  ++KN++   T++++++ VK+ P+L+LDDIGAE
Sbjct: 178 LAAIANELAEKE-YSSMIVYVPEFVRELKNSLQDQTLEEKLNMVKTTPVLMLDDIGAESM 236

Query: 229 TSWVRDEILQVILQHRMLEELPTFFTSNYSFNDLERKWA-NIKGSDETWQAKRVMERVRY 287
           TSWVRDE++  +LQHRM ++LPTFF+SN+S ++L+  +  + +G  E  +A R+MER+ Y
Sbjct: 237 TSWVRDEVIGTVLQHRMSQQLPTFFSSNFSPDELKHHFTYSQRGEKEEVKAARLMERILY 296

Query: 288 LAIEFHLEGPNRR 300
           LA    L+G NRR
```

```
Sbjct: 297 LAAPIRLDGENRR 309
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4833> which encodes the amino acid sequence <SEQ ID 4834>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1944(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 228/300 (76%), Positives = 264/300 (88%)

Query: 1   MKSVGQALENQGRVPRNTNDELIQMILADAQVAEFIKTHQLSQREINISMSKFNQFLIER 60
           M+ +G+ +   G+  R  +D+LIQ ILAD +VA FI  H LSQ +IN+S+SKFNQFL+ER
Sbjct: 1   MEKIGETMAKLGQNTRVNSDQLIQTILADPEVASFISQHHLSQEQINLSLSKFNQFLVER 60

Query: 61  QKFKNKDSQYIAKGYEPILVMNEGYADVSYLETRELIEAQKKQAISDRINLVNLPKSYRN 120
           QK++ KD  YIAKGY+PIL MNEGYADVSYLET+EL+EAQK+ AIS+RI LV+LPKSYR+
Sbjct: 61  QKYQLKDPSYIAKGYQPILAMNEGYADVSYLETKELVEAQKQAAISERIQLVSLPKSYRH 120

Query: 121 IRMTDFDINNESRMKAMSQLLDFVETYPSYNHKGLYLYGDMGVGKSYLMAAMARELSERK 180
           I ++D D+NN SRM+A S +LDFVE YPS   KGLYLYGDMG+GKSYL+AAMA ELSE+K
Sbjct: 121 IHLSDIDVNNASRMEAFSAILDFVEQYPSAEQKGLYLYGDMGIGKSYLLAAMAHELSEKK 180

Query: 181 GVSTTLLHFPSFAIDVKNAISSGTVKDEIDAVKSVPILILDDIGAEQATSWVRDEILQVI 240
           GVSTTLLHFPSFAIDVKNAIS+G+VK+EIDAVK+VP+LILDDIGAEQATSWVRDE+LQVI
Sbjct: 181 GVSTTLLHFPSFAIDVKNAISNGSVKEEIDAVKNVPVLILDDIGAEQATSWVRDEVLQVI 240

Query: 241 LQHRMLEELPTFFTSNYSFNDLERKWANIKGSDETWQAKRVMERVRYLAIEFHLEGPNRR 300
           LQ+RMLEELPTFFTSNYSF DLERKWA IKGSDETWQAKRVMERVRYLA EFHLEG NRR
Sbjct: 241 LQYRMLEELPTFFTSNYSFADLERKWATIKGSDETWQAKRVMERVRYLAREFHLEGANRR 300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1565

A DNA sequence (GBSx1658) was identified in *S.agalactiae* <SEQ ID 4835> which encodes the amino acid sequence <SEQ ID 4836>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2660(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4837> which encodes the amino acid sequence <SEQ ID 4838>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2135(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 217/391 (55%), Positives = 309/391 (78%)

Query: 1   MMSPIDEFTYIKQNKIVYDSNSLIQLYFPIMGSDAMALYDYFVHFFDDGIRRHKFSEVLN 60
           MM PID FTY+K+NK+  DS +LIQLYFPI+GSDA+++Y YF+HFFDDG++RHKFS++LN
Sbjct: 1   MMKPIDTFTYLKRNKVTLDSVTLIQLYFPIIGSDAVSIYQYFIHFFDDGLQRHKFSDILN 60

Query: 61  HLQYGMPRFQDALVMLTALDLLTVYQATGTYLVKLNQAMSNELFLSNPIYRRLLEKRIGE 120
           HLQ+GM RF+DAL +LTA++L++VYQ + TYL+ L+Q +S +LF  +P Y RLLE++IGE
Sbjct: 61  HLQFGMKRFEDALAILTAMELVSVYQLSDTYLITLHQPLSRDLFFQHPAYSRLLEQKIGE 120

Query: 121 VAVAELDMKIPKNARDISKKFTDVFSDLGQPKQEVNRSKNVFDLESFKRLMMRDGLRFNN 180
           VAV+EL + +P  AR+ISK+F+D+F   G       + +  FDL SF++LM+RDGL+F +
Sbjct: 121 VAVSELQVTVPSQARNISKRFSDIFGVQGDLTNVPQKPQKNFDLSSFQQLMVRDGLQFED 180

Query: 181 EKDDVLGIYSVSELYHLNWYDTYQLAKQTAINGMIAPQRMKVQQNEGQHIKDNQSFTNNE 240
            + D++ +YS++E Y + W+DTYQ+AK TA+NG I P+R+  ++N+       ++F+  E
Sbjct: 181 NQKDIISLYSIAEQYDMTWFDTYQIAKATAVNGKIRPERLLAKKNQSMTKPSKENFSQAE 240

Query: 241 KVILRESKNDSALVFLEKIKRSRKAVTTSGEKTLLEDLAKMNFLDEVINVMVLYTLNKTK 300
           ++ILRE+K DSALVFLEKIK++R+A  T  E+ LL+ LAKMNFLD+VINVMVLYT NKTK
Sbjct: 241 QIILREAKQDSALVFLEKIKKARRATITKDERILLQTLAKMNFLDDVINVMVLYTFNKTK 300

Query: 301 SANLNKAYIMKVANDFAFQNVMTAEDAVLKIRDFSDQKVRTKTETKKKQSNVPEWSNPDY 360
           SANL K+Y++K+ANDFA+Q V TAE+A++ +R F+D++ R +++ K  QSNVP+WSNPDY
Sbjct: 301 SANLQKSYVLKMANDFAYQKVSTAEEAIVVLRAFTDRQSRRQSKVKTSQSNVPKWSNPDY 360

Query: 361 KDEVSPEKEIELEQFKTDALKRLERLGKDGE 391
           ++  S E++ +L+QFK  ALKRLE LGK G+
Sbjct: 361 QETTSQEEQAKLDQFKQAALKRLENLGKGGD 391
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1566

A DNA sequence (GBSx1659) was identified in *S.agalactiae* <SEQ ID 4839> which encodes the amino acid sequence <SEQ ID 4840>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4485(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06865 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 80/150 (53%), Positives = 115/150 (76%)

Query: 1   MRCPKCGYNKSSVVDSRQAEEGTTIRRRRECEKCGNRFTTFERLEELPLLVIKKDGTREQ 60
           MRCP C +N + V+DSR A EG +IRRRRECE C +RFTTFE +EE+PL+V+KKDGTR++
Sbjct: 1   MRCPACHHNGTRVLDSRPAHEGRSIRRRRECESCNHRFTTFEMIEEVPLIVVKKDGTRQE 60

Query: 61  FSRDKILNGIIQSAQKRPVSSEDIENCILRIERKIRSEYEDEVSSITIGNLVMDELAELD 120
           FS DKIL G+I++ +KRPV  E +E  +  +ER++R + ++EV S  IG LVM+ LA +D
Sbjct: 61  FSSDKILRGLIRACEKRPVPLETLEGIVNEVERELRGQGKNEVDSKEIGELVMERLANVD 120

Query: 121 EITYVRFASVYKSFKDVDEIEELLQQITKR 150
           ++ YVRFASVY+ FKD++   + L+++ +R
Sbjct: 121 DVAYVRFASVYRQFKDINVFIQELKELMER 150
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4841> which encodes the amino acid sequence <SEQ ID 4842>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4365(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 131/155 (84%), Positives = 143/155 (91%)

Query: 1   MRCPKCGYNKSSVVDSRQAEEGTTIRRRRECEKCGNRFTTFERLEELPLLVIKKDGTREQ 60
            +RCPKC Y+KSSVVDSRQAE+G TIRRRRECE+C  RFTTFER+EELPLLVIKKDGTREQ
Sbjct: 1   VRCPKCNYHKSSVVDSRQAEDGNTIRRRRECEQCHTRFTTFERVEELPLLVIKKDGTREQ 60

Query: 61  FSRDKILNGIIQSAQKRPVSSEDIENCILRIERKIRSEYEDEVSSITIGNLVMDELAELD 120
            FSRDKILNG++QSAQKRPVSS DIEN I RIE+++R+ YE+EVSS  IGNLVMDELAELD
Sbjct: 61  FSRDKILNGVVQSAQKRPVSSTDIENVISRIEQEVRTTYENEVSSTAIGNLVMDELAELD 120

Query: 121 EITYVRFASVYKSFKDVDEIEELLQQITKRVRSKK 155
            EITYVRFASVYKSFKDVDEIEELLQQIT RVR KK
Sbjct: 121 EITYVRFASVYKSFKDVDEIEELLQQITNRVRGKK 155
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1567

A DNA sequence (GBSx1660) was identified in *S.agalactiae* <SEQ ID 4843> which encodes the amino acid sequence <SEQ ID 4844>. This protein is predicted to be CsrS (mtrB). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.30   Transmembrane   22 -  38 (  18 -  43)
   INTEGRAL    Likelihood = -9.66   Transmembrane  189 - 205 ( 187 - 212)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2109> which encodes the amino acid sequence <SEQ ID 2110>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.32   Transmembrane  196 - 212 ( 189 - 214)

----- Final Results -----
               bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 248/501 (49%), Positives = 363/501 (71%), Gaps = 4/501 (0%)
```

```
Query: 1   MKNKKDQFIGVKQPLSKKLSQLVFILFFSLFTVFSVLVYTSATRYVLHREKINVGRSLEK 60
           M+N+K +    K  L K+LS + F+LFF +F+ F+++ Y+S   ++L +EK +V +++
Sbjct: 1   MENQKQKQKKYKNSLPKRLSNIFFVLFFCIFSAFTLIAYSSTNYFLLKKEKQSVFQAVNI 60

Query: 61  TRVRLSQANSSLTSDDILEILYNQVFADDIYPHKRQNGIVRTGESIDSILYVNQEMTLYD 120
            RVRLS+ +S+ T +++ E+LY           ++ + ++R+   I + L  NQ++ +Y+
Sbjct: 61  VRVRLSEVDSNFTLENLAEVLYKNDKTHLRIDDRKGSRVIRSERDITNTLDANQDIYVYN 120

Query: 121 VNRKPVFST-LRTGMPTIGKSMGKVIISKVADM-EGFVGTKAIYSQKTGQLLGYVQIFYN 178
           ++++ +F+T      P +   +G+V    + D   GF  T+ +YS +TG+ +GYVQ+F++
Sbjct: 121 IDKQMIFTTDNEESSPGLHGPIGRVYHDHIEDQYRGFSMTQKVYSNRTGKFVGYVQVFHD 180

Query: 179 LGRYYSMRQNIIVFLIMMEVLGTVLALVVINSATKRIVRPVKNLHDLMHQISENPSNLEI 238
           LG YY +R  ++ +L+++E+ GT LA ++I   T+R ++P+ NLH++M  ISENP+NL +
Sbjct: 181 LGNYYVIRARLLFWLLVVELFGTSLAYLIILITTRRFLKPLHNLHEVMRNISENPNNLNL 240

Query: 239 RSKVRSEDEIGELSRIFDGMLDQLEDYTRRQSQFISDVSHELRTPVAVVKGHIGLLQRWG 298
           RS + S DEI ELS IFD MLD+LE +T+ QS+FISDVSHELRTPVA++KGHIGLLQRWG
Sbjct: 241 RSDISSGDEIEELSVIFDNMLDKLETHTKLQSRFISDVSHELRTPVAIIKGHIGLLQRWG 300

Query: 299 KDDPEILEESLAAAYHEADRMSLMINDMLNMIRVQGSLELHQDEVTDLSSSISVVIENFR 358
           KDD +ILEESL A  HEADRM++MINDML+MIRVQGS E HQ+++T L  SI  V+ NFR
Sbjct: 301 KDDSDILEESLTATAHEADRMAIMINDMLDMIRVQGSFEGHQNDMTVLEDSIETVVGNFR 360

Query: 359 ILREDFQFIFENNISDIVWGKIYKIHFEQALMILIDNAIKYSPSYKEVSVVLSVDNDFAT 418
           +LREDF F +++     +  +IYK HFEQALMILIDNA+KYS   K++++ LSV
Sbjct: 361 VLREDFIFTWQSENPKTI-ARIYKNHFEQALMILIDNAVKYSRKEKKIAINLSVTGKQEA 419

Query: 419 VV-VKDKGEGISDEDIEFIFDRFYRTDKSRNRESTQAGLGIGLSVFKQIMDAYHLKVDIK 477
           +V V+DKGEGIS EDIE IF+RFYRTDKSRNR STQAGLGIGLS+ KQI+D YHL++ ++
Sbjct: 420 IVRVQDKGEGISKEDIEHIFERFYRTDKSRNRTSTQAGLGIGLSILKQIVDGYHLQMKVE 479

Query: 478 SELNQGTEFIVRIPIKKFEET 498
           SELN+G+ FI+ IP+ + +E+
Sbjct: 480 SELNEGSVFILHIPLAQSKES 500
```

A related GBS gene <SEQ ID 8845> and protein <SEQ ID 8846> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
SRCFLG: 0
McG: Length  of  UR:   5
     Peak Value of UR:   0.74
     Net Charge of CR: 2
McG: Discrim Score:    -10.19
GvH: Signal Score (-7.5): -3.66
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value: -11.30 threshold:  0.0
   INTEGRAL    Likelihood =-11.30   Transmembrane   22 -  38 (  18 -  43)
   INTEGRAL    Likelihood = -9.66   Transmembrane  189 - 205 ( 187 - 212)
   PERIPHERAL  Likelihood =  2.86     405
 modified ALOM score:   2.76
icml HYPID: 7  CFP: 0.552

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8846 (GBS321) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 6; MW 84kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 2; MW 58.7kDa).

GBS321-GST was purified as shown in Figure 220, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1568

A DNA sequence (GBSx1661) was identified in *S.agalactiae* <SEQ ID 4845> which encodes the amino acid sequence <SEQ ID 4846>. This protein is predicted to be CsrR (trcR). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2649(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3259> which encodes the amino acid sequence <SEQ ID 3260>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3226(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 193/229 (84%), Positives = 211/229 (91%), Gaps = 1/229 (0%)

Query: 1   MGKKILIIEDEKNLARFVSLELLHEGYDVVVETNGREGLDTALEKDFDLILLDLMLPEMD 60
           M KKILIIEDEKNLARFVSLEL HEGY+V+VE NGREGL+TALEK+FDLILLDLMLPEMD
Sbjct: 1   MTKKILIIEDEKNLARFVSLELQHEGYEVIVEVNGREGLETALEKEFDLILLDLMLPEMD 60

Query: 61  GFEITRRLQAEKTTYIMMMTARDSVMDIVAGLDRGADDYIVKPFAIEELLARVRAIFRRQ 120
           GFE+TRRLQ EKTTYIMMMTARDS+MD+VAGLDRGADDYIVKPFAIEELLAR+RAIFRRQ
Sbjct: 61  GFEVTRRLQTEKTTYIMMMTARDSIMDVVAGLDRGADDYIVKPFAIEELLARIRAIFRRQ 120

Query: 121 EIETKTKEKGDSGSFRDLSLNTHNRSAMRGDEEISLTKREFDLLNVLMTNMNRVMTREEL 180
           +IE++ K+    G +RDL LN  NRS  RGD+EISLTKRE+DLLN+LMTNMNRVMTREEL
Sbjct: 121 DIESE-KKVPSQGIYRDLVLNPQNRSVNRGDDEISLTKREYDLLNILMTNMNRVMTREEL 179

Query: 181 LEHVWKYDVAAETNVVDVYIRYLRGKIDIPGRESYIQTVRGMGYVIREK 229
           L +VWKYD A ETNVVDVYIRYLRGKIDIPG+ESYIQTVRGMGYVIREK
Sbjct: 180 LSNVWKYDEAVETNVVDVYIRYLRGKIDIPGKESYIQTVRGMGYVIREK 228
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1569

A DNA sequence (GBSx1662) was identified in *S.agalactiae* <SEQ ID 4847> which encodes the amino acid sequence <SEQ ID 4848>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3864(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG32547 GB:U12643 YlbN-like hypothetical protein [Streptococcus gordonii]
 Identities = 91/174 (52%), Positives = 133/174 (76%), Gaps = 3/174 (1%)

Query: 3   LTEIKKSPEGLYFDKKIDIKESLMERHSEIMDISDIQVSGHVVYEDGLYLLDYNMAYDIT 62
           + EI+K+P+GL F+KK+D+ E L ER++EI+D+ DI  SG   YEDGLY LDY ++Y IT
Sbjct: 4   IQEIRKNPDGLAFEKKLDLAEELKERNAEILDVQDIVASGRAQYEDGLYFLDYELSYTIT 63

Query: 63  LPSSRSMKPVVLSEKQTINEVFIEAENVSTKKELVDQELVLILEEDDINLEESVIDNILL 122
           L SSRSM+PV   E   +NE+F+E   V++ +E++DQ+LVL +E  +IN+ ESV DNILL
Sbjct: 64  LASSRSMEPVERKESYLVNEIFMEDGQVAS-QEMIDQDLVLPIENGEINVAESVADNILL 122

Query: 123 NIPLRVL-AADEVGVEADLSGKNWSLMTEKQYEEKQAKEKEKSNPFAALEGMFD 175
           NIPL+VL AA+E G +   +G++W +MTE  Y++ QA++KE+++PFA L+G+FD
Sbjct: 123 NIPLKVLTAAEEAGSDLP-TGRDWQVMTEDDYQKYQAEKKEENSPFAGLQGLFD 175
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4849> which encodes the amino acid sequence <SEQ ID 4850>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3032(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 86/175 (49%), Positives = 135/175 (77%)

Query: 1   MLLTEIKKSPEGLYFDKKIDIKESLMERHSEIMDISDIQVSGHVVYEDGLYLLDYNMAYD 60
           + ++EI+K P+GL FD+  D+K  L+ER  +I+DI  ++  G+V Y+ GLYLLDY ++Y+
Sbjct: 3   LAISEIRKHPDGLSFDRLCDVKSMLLERDQQIIDIKAVKAVGNVRYDKGLYLLDYQLSYE 62

Query: 61  ITLPSSRSMKPVVLSEKQTINEVFIEAENVSTKKELVDQELVLILEEDDINLEESVIDNI 120
           + LPSSRSM PV LSE Q I E+FIEA +++ KKELV+  LVL+L++D INLEES++DNI
Sbjct: 63  VILPSSRSMVPVCLSEVQHIQELFIEATDLADKKELVEDNLVLVLDKDAINLEESIVDNI 122

Query: 121 LLNIPLRVLAADEVGVEADLSGKNWSLMTEKQYEEKQAKEKEKSNPFAALEGMFD 175
           LL IP++VL  +E   +   +G+NW+++TE+ Y+  + ++++++NPFA+L+G+FD
Sbjct: 123 LLAIPVQVLTEEEKKSKELPAGQNWAVLTEEDYQCLKEEKQKENNPFASLQGLFD 177
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1570

A DNA sequence (GBSx1663) was identified in *S.agalactiae* <SEQ ID 4851> which encodes the amino acid sequence <SEQ ID 4852>. This protein is predicted to be heat shock protein (htpX). Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.30   Transmembrane  195 - 211 ( 190 - 221)
```

```
  INTEGRAL    Likelihood =-11.09   Transmembrane   43 -  59 (  31 -  62)
  INTEGRAL    Likelihood = -3.61   Transmembrane  153 - 169 ( 153 - 174)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB70525 GB:AF017421 putative heat shock protein HtpX
            [Streptococcus gordonii]
 Identities = 220/297 (74%), Positives = 261/297 (87%), Gaps = 1/297 (0%)

Query: 1   MLYQQIASNKRKTVVLLIVFFCLLAAIGAAVGYLVLGSYQFGLVLALIIGVIYAVSMIFQ 60
            ML++QIA+NKR+T  LL+ FF LLA IGAA GYL + S   G+++A IIG+IYA++MIFQ
Sbjct: 1   MLFEQIAANKRRTWFLLVAFFALLALIGAAAGYLWMNSPLGGVIIAFIIGLIYAITMIFQ 60

Query: 61  STNVVMSMNNAREVTEDEAPNYFHIVEDMAMIAQIPMPRVFIVEDDSLNAFATGSKPENA 120
            ST VVMSMN AR+V+E EAP  +HIV+DMAM+AQIPMPRV+IVEDDS NAFATGS PENA
Sbjct: 61  STEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPMPRVYIVEDDSPNAFATGSNPENA 120

Query: 121 AVAATTGLLAVMNREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRMLFYG 180
            AVAATTGLL +MNREELEGVIGHEVSHIRNYDIRISTIAVALASA+T+ISS+  RM++YG
Sbjct: 121 AVAATTGLLRLMNREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMMWYG 180

Query: 181 GGRRRDDDREDGG-NILVLIFSILSLILAPLAASLVQLAISRQREYLADASSVELTRNPQ 239
            GGRRR+D  +D G  +L+L+FS++++ILAPLAA+LVQLAISRQRE+LADASSVELTRNPQ
Sbjct: 181 GGRRRNDRDDDSGLGLLMLVFSLIAIILAPLAATLVQLAISRQREFLADASSVELTRNPQ 240

Query: 240 GMISALEKLDRSEPMGHPVDDASAALYINDPTKKEGLKSLFYTHPPIADRIERLRHM 296
            GMI AL+KLD SEPM   VDDASAALYI+DP KK GL+ LFYTHPPI++R+ERLR M
Sbjct: 241 GMIRALQKLDNSEPMHRHVDDASAALYISDPKKKGGLQKLFYTHPPISERVERLRKM 297
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4853> which encodes the amino acid sequence <SEQ ID 4854>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31
>>> Seems to have a cleavable N-term signal seq.
  INTEGRAL    Likelihood = -9.77   Transmembrane  197 - 213 ( 192 - 223)
  INTEGRAL    Likelihood = -8.33   Transmembrane   43 -  59 (  33 -  61)
  INTEGRAL    Likelihood = -3.82   Transmembrane  153 - 169 ( 153 - 174)

----- Final Results -----
               bacterial membrane --- Certainty=0.4906(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB70525 GB:AF017421 putative heat shock protein HtpX [Streptococcus gordonii]
 Identities = 208/298 (69%), Positives = 257/298 (85%), Gaps = 1/298 (0%)

Query: 1   MLYQQISQNKQRTVVLLVGFFALLALIGASAGYLLLDNYAMGLVLALVIGVIYATSMIFQ 60
            ML++QI+ NK+RT  LLV FFALLALIGA+AGYL +++   G+++A +IG+IYA +MIFQ
Sbjct: 1   MLFEQIAANKRRTWFLLVAFFALLALIGAAAGYLWMNSPLGGVIIAFIIGLIYAITMIFQ 60

Query: 61  STSLVMSMNNAREVTEKEAPGFFHIVEDMAMVAQIPMPRVFIIEDPSLNAFATGSSPQNA 120
            ST +VMSMN AR+V+E+EAP  +HIV+DMAMVAQIPMPRV+I+ED S NAFATGS+P+NA
Sbjct: 61  STEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPMPRVYIVEDDSPNAFATGSNPENA 120

Query: 121 AVAATTGLLEVMNREELEGVIGHEISHIRNYDIRISTIAVALASAVTVISSIGGRMLWYG 180
            AVAATTGLL +MNREELEGVIGHE+SHIRNYDIRISTIAVALASA+T+ISS+ GRM+WYG
Sbjct: 121 AVAATTGLLRLMNREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMMWYG 180

Query: 181 GGSRRQRDDGDDDVLRIITLLLSLLSLLLAPLVASLIQLAISRQREYLADASSVELTRNP 240
            GG RR+ D  DD  L ++ L+ SL++++LAPL A+L+QLAISRQRE+LADASSVELTRNP
```

```
Sbjct: 181 GG-RRRNDRDDDSGLGLLMLVFSLIAIILAPLAATLVQLAISRQREFLADASSVELTRNP 239

Query: 241 QGMIKALEKLQLSQPMKHPVDDASAALYINEPRKKRSFSSLFSTHPPIEERIERLKNM 298
           QGMI+AL+KL  S+PM   VDDASAALYI++P+KK     LF THPPI ER+ERL+ M
Sbjct: 240 QGMIRALQKLDNSEPMHRHVDDASAALYISDPKKKGGLQKLFYTHPPISERVERLRKM 297
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 233/298 (78%), Positives = 262/298 (87%), Gaps = 2/298 (0%)

Query: 1   MLYQQIASNKRKTVVLLIVFFCLLAAIGAAVGYLVLGSYQFGLVLALIIGVIYAVSMIFQ 60
           MLYQQI+ NK++TVVLL+ FF LLA IGA+ GYL+L +Y  GLVLAL+IGVIYA SMIFQ
Sbjct: 1   MLYQQISQNKQRTVVLLVGFFALLALIGASAGYLLLDNYAMGLVLALVIGVIYATSMIFQ 60

Query: 61  STNVVMSMNNAREVTEDEAPNYFHIVEDMAMIAQIPMPRVFIVEDDSLNAFATGSKPENA 120
           ST++VMSMNNAREVTE EAP +FHIVEDMAM+AQIPMPRVFI+ED SLNAFATGS P+NA
Sbjct: 61  STSLVMSMNNAREVTEKEAPGFFHIVEDMAMVAQIPMPRVFIIEDPSLNAFATGSSPQNA 120

Query: 121 AVAATTGLLAVMNREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRMLFYG 180
           AVAATTGLL VMNREELEGVIGHE+SHIRNYDIRISTIAVALASAVT+ISSIG RML+YG
Sbjct: 121 AVAATTGLLEVMNREELEGVIGHEISHIRNYDIRISTIAVALASAVTVISSIGGRMLWYG 180

Query: 181 GG--RRRDDDREDGGNILVLIFSILSLILAPLAASLVQLAISRQREYLADASSVELTRNP 238
           GG  R+RDD  +D   I+ L+ S+LSL+LAPL ASL+QLAISRQREYLADASSVELTRNP
Sbjct: 181 GGSRRQRDDGDDDVLRIITLLLSLLSLLLAPLVASLIQLAISRQREYLADASSVELTRNP 240

Query: 239 QGMISALEKLDRSEPMGHPVDDASAALYINDPTKKEGLKSLFYTHPPIADRIERLRHM 296
           QGMI ALEKL  S+PM HPVDDASAALYIN+P KK    SLF THPPI +RIERL++M
Sbjct: 241 QGMIKALEKLQLSQPMKHPVDDASAALYINEPRKKRSFSSLFSTHPPIEERIERLKNM 298
```

A related GBS gene <SEQ ID 8847> and protein <SEQ ID 8848> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      9.61
GvH: Signal Score (-7.5): -0.97
     Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
ALOM program  count: 3 value: -11.30 threshold:  0.0
   INTEGRAL    Likelihood =-11.30   Transmembrane  195 - 211 ( 190 - 221)
   INTEGRAL    Likelihood =-11.09   Transmembrane   43 -  59 (  31 -  62)
   INTEGRAL    Likelihood = -3.61   Transmembrane  153 - 169 ( 153 - 174)
   PERIPHERAL  Likelihood =  5.89      87
 modified ALOM score:    2.76

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
73.8/88.3% over 296aa
                                                                imported
  SP|O30795| PUTATIVE HEAT SHOCK PROTEIN HTPX. Insert characterized
  GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus
gordonii} Insert characterized
  PIR|T48855|T48855 probable heat shock protein HtpX - Streptococcus gordonii Insert
characterized

ORF02338(301 - 1188 of 1488)
SP|O30795|HTPX_STRGC(1    -    297    of    297)   PUTATIVE   HEAT   SHOCK   PROTEIN
HTPX.GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus
gordonii}PIR|T48855|T48855 probable heat shock protein HtpX [imported] - Streptococcus
gordonii
```

```
%Match = 44.0
%Identity = 73.7  %Similarity = 88.2
Matches = 219  Mismatches = 34  Conservative Sub.s = 43

141       171       201       231       261       291       321       351
NFLFTSVI*HNNIQL*CEIRNFPK*YCWKTIWVQTKPILRNS*RRKRSAKSFL*LLIEKGERLLLYQQIASNKRKTVVLL
                                                               :|::|||:|||:|  ||
                                                               MLFEQIAANKRRTWFLL
                                                                       10

381       411       441       471       501       531       561       591
IVFFCLLAAIGAAVGYLVLGSYQFGLVLALIIGVIYAVSMIFQSTNVVMSMNNAREVTEDEAPNYFHIVEDMAMIAQIPM
: || ||| |||| ||| : |   |:::|:|||:|||::|||||| |||||| ||:|:| |||  :|||:||||:|||||
VAFFALLALIGAAAGYLWMNSPLGGVIIAFIIGLIYAITMIFQSTEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPM
           30        40        50        60        70        80        90

621       651       681       711       741       771       801       831
PRVFIVEDDSLNAFATGSKPENAAVAATTGLLAVMNREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRML
|||:|||||| ||||||| |||||||||||||| :||||||||||||||||||||||||||||||||:|:|||:  ||:
PRVYIVEDDSPNAFATGSNPENAAVAATTGLLRLMNREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMM
          110       120       130       140       150       160       170

861       888       918       948       978      1008      1038      1068
FYGGGRRRDDDREDGG-NILVLIFSILSLILAPLAASLVQLAISRQREYLADASSVELTRNPQGMISALEKLDRSEPMGH
:|||||||:|  :| |  :|:|:||::::||||||:|||||||||||:||||||||||||||||| ||:||| ||||
WYGGGRRRNDRDDDSGLGLLMLVFSLIAIILAPLAATLVQLAISRQREFLADASSVELTRNPQGMIRALQKLDNSEPMHR
          190       200       210       220       230       240       250

1098      1128      1158      1188      1218      1248      1278      1308
PVDDASAALYINDPTKKEGLKSLFYTHPPIADRIERLRHM*SLTKRRVAMPCVLFF*DKACKT*YNMTYTIKGDGTCYLQ
 ||||||||||:|| || ||: ||||||||::|:|||| |
HVDDASAALYISDPKKKGGLQKLFYTHPPISERVERLRKM
          270       280       290
```

SEQ ID 8848 (GBS179) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 11; MW 58kDa).

GBS179-GST was purified as shown in Figure 227, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1571

A DNA sequence (GBSx1665) was identified in *S.agalactiae* <SEQ ID 4855> which encodes the amino acid sequence <SEQ ID 4856>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-15.44   Transmembrane    4 -  20 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG23700 GB:AF017421 LemA-like protein [Streptococcus gordonii]
 Identities = 124/182 (68%), Positives = 152/182 (83%)

Query: 1    MGTMILIAIIALFVIWLIVAYNSLVRSRMHTKESWSQIDVQLKRRNDLIPNLIETVKGYA 60
            M  +I IA+I + V+++I  YNSLVR+RM T+E+WSQIDVQLKRRNDL+PNLIETVKGY
Sbjct: 1    MSFIITIAVIVVIVLFVISVYNSLVRARMQTQEAWSQIDVQLKRRNDLLPNLIETVKGYG 60

Query: 61   AYEGKTLEKIAELRAQVAKANTPAEAMTASNELTRQLSSILAVAENYPDLKANNSFVKLQ 120
```